

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 15:48:35 ; Search time 6225.59 Seconds
(without alignments)
14215.251 Million cell updates/sec

Title: US-09-316-163-1

Perfect score: 4229

Sequence: 1 tcgagtcgaactgtcccgag.....aaaaaaaaaaaaaaaaaaaa 4229

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_ov:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_inv:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

ALIGNMENTS

RESULT 1

A91883

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

A91883

Sequence 1 from Patent WO9823638.

A91883

A91883.1

GI:6740760

unidentified.

unclassified.

1 (bases 1 to 4229)

Schwaebler, W. and Slin, R. B.

COMPLEMENT INHIBITOR

Patent: WO 9823638-A 1 04-JUN-1998;

SCHWAEBLER WILHELM (GB); UNIV LEICESTER (GB)

Location/Qualifiers

1..4229

/organism="unidentified"

/db_xref="taxon:32644"

1395 a

791 c

893 g

1150 t

DNA

linear

PAT 22-JAN-2000

A91883 Sequence 1
AJ30522 Rattus no
M12660 Mouse Cfh 1
A91885 Sequence 3
Y00716 Human mRNA
M29009 Mouse compl
AF436847 Rattus no
A91886 Sequence 4
M17517 Human compl
X98697 B.taurus mr
X07523 Human mRNA
X04697 Human mRNA
A91884 Sequence 2
M29010 Mouse compl
BC012610 Homo sapi
AF348145 Rattus no
M29007 Mouse compl
AX274975 Sequence
M3274975 Sequence
AF295327 Homo sapi
AX274977 Sequence
M12383 Human compl
AC096417 Rattus no
AF348144 Rattus no
M29008 Mouse compl
M65293 Human fact
BC016755 Homo sapi
X56209 Human mRNA
X56210 H.sapiens m
M65292 Human fact
M65294 Human fact
X68679 H.sapiens
BC020687 Homo sapi
M23554 Mouse compl
M31980 Mouse fact
AC096417 Rattus no
X98337 H.sapiens m
BC022283 Homo sapi
X64877 H.sapiens m
AC097078 Rattus no
AC100698 Mus muscu
AC097078 Rattus no
AC100698 Mus muscu
AL049744 Human DNA
AL161735 Homo sapi
AC026056 Homo sapi

Query Match		100.0%	Score 4229;	DB 6;	Length 4229;		
Best Local Similarity		100.0%	Pred. No. 0;	Mismatches	0;	Indels	0;
Matches 4229;		Conservative	0;				Gaps
Qy	1	tcgagtcactgctccagatagatccagacatgagactgacagagatgtaaaaggtctctcccaagagaaaatt	120				
Db	1	TCGAGTCAACTGCTCCAGATAGATCCAGACATGAGACTGTCCACGAGAATATTGGC	60				
Qy	61	ttatattatggaactgttgtagcagagaattgttaaaggtctctcccaagagaaaatt	120				
Db	61	TTATATTATGGAAGTGTGTGTGTAGCAGAGAATGTGTAAGAGTCTCTCCCAAGAGAAAATT	120				
Qy	121	cagaatctctcaggttcggtgctgaactatattcagaagggcactcagcaacct	180				
Db	121	CAGAAATTTCTCAGCTTCGGTGTGACAACTATATTCAGAGGCACTCAGGCAACCT	180				
Qy	181	acaaatgcgccttgatccaggaacacttggtactattgtaaaagtacgaagaatggag	240				
Db	181	ACAAATGCCGCCCTGATACCAACCACTTGGTACTATTGTAAGATATGCAAGAAATGGAG	240				
Qy	241	aatgggtacctcttaaccatcaagatatagtcggaagggccatgtgggcatcccgag	300				
Db	241	AATGGGTACCTTCTAACCCATCAAGGATATGTGGAAAAGGCCATGTGGGCATCCCGGAG	300				
Qy	301	acacaccccttgggtccctttaggtcggcagttggatctggaatttgaatttggtgcaagg	360				
Db	301	ACACACCCCTTGGGTCCCTTTAGGCTGGCAGTTGGATCTGAAATTTGGTGCAAAAGG	360				
Qy	361	ttgtttatcatgtgatgaagggtaccacactattaggtgaaattgattaccgtgaatgtg	420				
Db	361	TTGTTTATACATGTGATCAAGGGTACCAACTATTAGGTGAAATTCATTCGCTGAATGTG	420				
Qy	421	atgcagatgggtggaccaatgatattccaatatgtgaagttgtaagttgctgccaagtg	480				
Db	421	ATGCAGATGGGTGGACCAATGATATTCCAATATGTGAAGTTGTGAAGTGTGCGCAGTGA	480				
Qy	481	cagaactggagaatggaagaattgtgagtggtgagcggacccagaccaggaattatt	540				
Db	481	CAGAACTGGAATGGAAGAATTGTGAGTGGTGCAGCCGCAACCAAGGGAATATTATT	540				
Qy	541	ttggaacaggttgtaactccggtcccaagattgaaaggacagaaagaaa	600				
Db	541	TTGGACAGGTGGTACGCTTTGAAATGCAACTCCGGCTTCAAGATTGAAGCAGAGAAAGAA	600				
Qy	601	tgcactgctataaaatggcctctggagcaatgaaagccacagtggtgaaatttctt	660				
Db	601	TGCAGTCTATAAATGGCCTCTGGACAAATGAAAGCCACAGTGTGTGAAATTTCTT	660				
Qy	661	gcctgccaccacgagttgaaaatggagatgatatctgaaaccagtttacaaaggag	720				
Db	661	GCCTGCCACCACGAGTTGAAAATGGAGATGCTATATCTGAACCAAGTTTACAAAGGAGA	720				
Qy	721	atgaagattccaatataaatgtaagcaaggttttgtgtacaaagaaagggatgctg	780				
Db	721	ATGAAGATTCCAAATATAAATGTAGCAAGGTTTGTGTACAAAAGAAAGGGGATGCTG	780				
Qy	781	tctgcaggggtctgagtggaatcctcagcctctctgtgaaagaatgacatgtttgactc	840				
Db	781	TCTGCAGGGGTCTGGATGGAAATCCTCAGCCCTCTGTGTAAGAAATGACATGTTTGACTC	840				
Qy	841	catatttccaaatgggtatctacacactcacaggattaaacacagaaattgatgataa	900				
Db	841	CATATTTCAAAATGGTATCTACACACTCACAGGATTAAACACAGAAATGTGATGATGAAA	900				
Qy	901	tcagatatgaatgtaaaatggcttctctctcagcccgatcacctgtttccaagtgt	960				
Db	901	TCAGATATGAATGTAAAAATGGCTTCTATCTCGTCAACCCGATCACCTGTTTCAAGTGTA	960				
Qy	961	caattactgctggaatccctctcgaagatgtagcttgaacacttgattttccacaat	1020				
Db	961	CAATTACTGCTGGATCCCTGCTCCAAGATGTAGCTTGAACCTTGTGATTTTCCCAAT	1020				
Qy	1021	tcgaacatggagctgtgtattatgaagaagcggagacccttctccagtagacctatag	1080				

Db	1021	TCAAAATGGAGCTGTGTATTATGAAGAAAGCCGGAGACCTTACTTCCAGTACCTATAG	1080				
Qy	1081	gaagagagacagctataactgtgacaacgggtttcacacgcttcacagtcatactggg	1140				
Db	1081	GAAAGGAGTACAGCTATAACTGTGACACGGGTTTACAACGGCTTTCACAGTCACTGGG	1140				
Qy	1141	actaccttcggtgcagctaaatgggtggagcctggaagttccatgcctcaggaatgta	1200				
Db	1141	ACTACTCTCGTTGGCAGCTAAATGGTGGAGCCTGAAGTTCCTCCCTCAGGCAATGTA	1200				
Qy	1201	ttttccattatgtgaatatggaatcttcactctgcacaaagaatatatagaggttc	1260				
Db	1201	TTTTCCATTATGTGGAATATGGAGAAATCTTCACTATGCGCAAGAGATATATAGAGGTC	1260				
Qy	1261	agctgcacaaagtcaggtgcagtgccagtgagctatagctctccaaatggtcgaagatacat	1320				
Db	1261	AGCTGCAAAAGTCCAGTCTCACAGTGGCTATAGTCTTCCAATGGTCAAGATACATATT	1320				
Qy	1321	attgtacagagaatggctgggtccctccctcccaaatgctccgatatcaagactgttccag	1380				
Db	1321	ATTGTACAGAGAATGGCTGGTCCCTCCCTCCCAAAATGGTCCGTATCAAGACTTGTTCAG	1380				
Qy	1381	tatcagatatagaaattgaaatgggttttttctcgaatcgtatcatatcatatgctctaa	1440				
Db	1381	TATCAGATATAGAAATGAAATGGGTTTTTTCTGAAATCTGATTTATACATATGCTCTAA	1440				
Qy	1441	atagaaaaacacggttatagatgtaaacaggatatgttaacaaatcccgaggaatctcag	1500				
Db	1441	ATAGAAAACACGGTATAGATGTAAACAGGATATGTAAACAATACCGGAGAAATATCAG	1500				
Qy	1501	gaataattacttctcgaatggatgggtcacctcgacctcatgcatctaaagttctgtg	1560				
Db	1501	GAATAATTACTTGTCTTCAAGATGGATGTCACCTCGACCTCATGCTTAAGTCTTGTG	1560				
Qy	1561	atatgcctgtatttgagaattctatgactaagaataataaacacatgggtttaaactcaatg	1620				
Db	1561	ATATGCCGTATTGAGAAATCTATGACTAAGAAATTAATTAACACATGGTTTAACTCAATG	1620				
Qy	1621	acaaattagactatgaatgtcacattggatatgaaatgaatataaaacataccaaagct	1680				
Db	1621	ACAAATTAGACTATGAATGTACATGGATGATGAAATGAATATTAACATACCAAGGCT	1680				
Qy	1681	ctataacatgtacttatgtatggatgggtctagtaacacctcctgttatgaaagagaatgca	1740				
Db	1681	CTATAACATGTACTTATGATGGTGTAGTACACCTCTGTTATGAAGAGAAATGCA	1740				
Qy	1741	gcattccctcttacacacacacttagttgttttccacagagaagtaaaatacaaatgtg	1800				
Db	1741	GCAITCCCTCTTACACCAAGACTTAGTTGTTTTTCCAGAGAGTAAATATACAAAGTTG	1800				
Qy	1801	gagattcgttgagttctcttgcggttcagacacagagttggagcagatttagtgcatt	1860				
Db	1801	GAGATTGTTGAGTTTCTCTTGGCTTCAGGACACAGAGTTGGAGCAGATTTAGTGCAAT	1860				
Qy	1861	gctaccactttggatgggtcccttaatttcccaacggtgtgaaggccaagtataaatctatg	1920				
Db	1861	GCTACCACCTTTGGATGGTCCCTTAATTTCCCAACGTGTGAAGGCCAAGTAAATCATCTG	1920				
Qy	1921	accaaactcttgaatcccgaaatggggaaataaaaggggaacaaaaaagttgaatacagcc	1980				
Db	1921	ACCAACCTCTTGAATCCCGAATCCGGAATGGGAAATTAAGGGGAACAAAAGTTGAATACAGCC	1980				
Qy	1981	atggtgacgtggtggaatatgattgcaaaccttagattctactgaaggagcccaataaaa	2040				
Db	1981	ATGGTGACGTGTTGGAATATGATTTGCAAACTAGATTCTACTGAAGGACCCCAATAAAA	2040				
Qy	2041	tccagttgttgacgggaagtggaacaggttgcgatatgctggttgatgagagaacat	2100				
Db	2041	TCCAGTGTGTGACGGGAAGTGGACAAAGTTGCCGATATGCGTTTGAGTATGAGAGAAT	2100				
Qy	2101	gtggagacccctctcgaactgagcatggctgtcgaagttatctgtccctccctaccatc	2160				

Db 2101 GTGGAGACCTCTCTGAACCTTGAGCATGGCTCTGCTCAAGTTATCTGTCTCCCTCCCTACCATC 2160
QY 2161 atggagattcagtgagggttcacttgcacagaaacaccttcacaatgattgacatgcagtag 2220
Db 2161 ATGGAGATTCACTGGAGCTTCACTTGTACAGAAACCTTCACAATGATTGCACATGCAGTAG 2220
QY 2221 tttctgcatttagtggaagtgagccagacttccctaattgtgttgcaacagatcaactg 2280
Db 2221 TTTTCTGCAATTAGTGGAAAGTGGACCGAGCTTCTCTCAATGTGTGGCAACAGATCAACTGG 2280
QY 2281 agaagtgtaaagcccggaagtcaactggcatagatgacaattccacaaataaauaatgaat 2340
Db 2281 ACAAGTGTAAAGCCCGGAAGTCAACTGGCATAGATCCAATTCATCCAANTAGAATGAAT 2340
QY 2341 ttaatacaactttagtgtgagtacagatgtagacaaagcagggtatgaacattcaaa 2400
Db 2341 TTAATCATAACTTTAGTGTCAAGTTACAGATGTAGACAAAAGCAGGAGTATGAACATTCAA 2400
QY 2401 tclgcatcaatggaagatggatccctgaaaccaaactgtcaagcaaaatclctgcaccl 2460
Db 2401 TCTGCAATCAATGGAAGTGGGATCTTGAAACCAACTGTACAAGCAAAAGATTCGCCCCC 2460
QY 2461 clccccgcagattccaaatgccaaagtgtatgaaacccacccgtgaaatacttgcaggg 2520
Db 2461 CTCCCCCGCAGATTCCAAATGCCAAAGTGATTGCAAAACCACCGTGAAATACTTCGATGGAG 2520
QY 2521 aaaaagtalcttcttcccaagatggttaccactacacacccagggcccaagaagaatgggt 2580
Db 2521 AAAAACTATCTCTGTCTTCTGCCAAGATGGTTACTTAACCTCAGGCGCCACGAAGAAATGCTGT 2580
QY 2581 gtaacatggaagggtggcagtcgttaccacgtgcacggaaaaaatcccatctcccaac 2640
Db 2581 GTAAACATGCAAGGTGGCAGTCTTACCACGCTGCACGGAAGAAATATCCATGTTCCAGC 2640
QY 2641 cccctaaalgtgaactggatctattaaatgcgccaggctccctcagaaagagaagatcl 2700
Db 2641 CCCCTAAAATTTCAACTGGCATCTATTAACTGCGCCAGGTCCTCAGAAAGAGAGCAGAGATT 2700
QY 2701 taatggatccagcagttatgaacggaactacattccagctattgctgtagaatggat 2760
Db 2701 TAAATGAGTCCAGCAGTTATGAACCGGAACCTACATPTCAGCTATTGCTCTAGAGATGGAT 2760
QY 2761 tcaagatatctgaaagaaatagggttaacnctgcaacalggaaaaatggagctclctgccl 2820
Db 2761 TCAAGATATCTGAAGAAATAGGGTAACCTGCAACATGGGAAATGGAGCTCTCTGCCCTC 2820
QY 2821 gttgtgtggaatacctgttggaacccccaccttcaattcccttgatgttcttcaltg 2880
Db 2821 GTTGTGTTGGAATACCTTGTGGACGCCACCTTCAATTCCTCTTGTGTAATTGTTCTCATC 2880
QY 2881 aactagaagttaccaatatggagaggaggttacatacaattgtrctgaaagctttggaa 2940
Db 2881 AACTAGAAAGTTACCAATATGGAGAGGAGGTTACATACAAATTTCTCTGAAGGCTTTTGGAA 2940
QY 2941 ttgatggaccagcatttataaagtgtaggaggacagtggtctgaarctccraaatgca 3000
Db 2941 TTGATGGACCACCAITTTATTAATGTGTAGGAGGACAGTGGTCTGNACCTCCCAATGCA 3000
QY 3001 taaaactgattgtgaacactgcccacatttgaattggcacaacccagacagaaaaa 3060
Db 3001 TAAAAACTGATTGTGACAACTTTGCCACATTTGAAATTTGCCAAACCGACAGAAAAAGAAA 3060
QY 3061 aaaaatcacaggttcaggagaacagtgacattcagatgtccactcngtatcgaaatgg 3120
Db 3061 AAAAAATCATACAGGTTCAGAGAACAGGTGACATTTCAGATGTCCAGCTCCGTATCGAATGG 3120
QY 3121 atggcctgacattgtcacatgtgttaataccgaaagtggattggacagccgggtatgcaag 3180
Db 3121 ATGGCTCTGACATTGTACATGTGTTAATACCAAGTGGATTGGACAGCGCGGTATGCAAG 3180
QY 3181 ataattcctgtgtgaatccaccacatgtgcgcaaatgctactactaatacaagggcaaga 3240
Db 3181 ATAATTCTGTGTGAATCCACCACATGTGCCAAATGCTACTATATACTAACAAGGCACAAGA 3240

QY 3241 ctaaatatcccatctggtgacaaagtacgttatgactgtaataaaccttttgaattattly 3300
Db 3241 CTAAATATCCATCTGGTGCACAAAGTACGTTATGACTGTAATAAACCTTTTGAATTATTG 3300
QY 3301 ggaagtggaaagtgatgtgccaaaaacgggaatttggacagaaacccaccgaaatgcaaaagalt 3360
Db 3301 GGAAGTGGAAAGTGATGTGCCAAAACGGGATTTGGACAGAACCCACCGAAATGCAAAAGATT 3360
QY 3361 caacaggggaatgtgggcctccctccacttatgtacaatggagacatcacctccttgcatt 3420
Db 3361 CAACAGGGAATGTGGGCTCTCCACCTATTGACAATGGAGACATCACCTCTCTGTGCAT 3420
QY 3421 taccagtatatgcaccattatcatcagltgaaatatcaatgccagaactattatclactla 3480
Db 3421 TACCAGTATATGCACCATTTATCATCAGTTCAATATCAATGCCAGAACATATTATCTACTTA 3480
QY 3481 agggaaataagatagtaacalglagaaaatggaaagtggtctcagcccaacacctgcttac 3540
Db 3481 AGGGAATTAAGATAGTAACATGTAGAAATGGAAAGTGGTCTCAGCCACCACCTGCTTAC 3540
QY 3541 atgcattgtgtataccagaagatattatggaaaaacataatatagttctcagatggaggy 3600
Db 3541 ATGCATGTGTGATACCCAGAAGATATTATGGAAAAACATAATATAGTTCTCAGATGGAGG 3600
QY 3601 aaaaatgcaagatttatcccaaaccaggggagaataltggaaatlcatgtglaaaacctggat 3660
Db 3601 AAAATGCAAGATTTATTCCCAATCAGGGGAGAAATATTGAATTCATGCTGTAAACCTGCAT 3660
QY 3661 atagaaaatccagagatcaccttccclctggtacaaaagtgcattggagggtcacatcaall 3720
Db 3661 ATAGAAAATTCAGAGGATCACCTCCCTTTGGTACAAAGTGCATTGAGGGTTCACATCAATT 3720
QY 3721 atcccactgtgtataaaatcgctalacaattatagtaaaaccttatggatggaaatgc 3780
Db 3721 ATCCCACCTGTGTATTAATCGCTATACAAATATTAGTAACCTTTAGGATGAGAAATGC 3780
QY 3781 acatgtatatactaaalacaglttgaattacatttaaatattgttttagctcatctcccl 3840
Db 3781 ACATGTATATTACTAATACAGTTGAATTTACATTTTAAATATTGTTAGCTCATCTTCCTC 3840
QY 3841 taataagtatataaaccttttttatatagtggttaatcagtaactttacagactgttggcc 3900
Db 3841 TAAATAAGTATATAAACCTTTTTTATATGGTGGTAAATCAGTAACCTTTACAGACTGTTGCC 3900
QY 3901 acaaaagcaagaacattcacattcaaaactcctaataccaaatatgatgtgtccaaaggacaaa 3960
Db 3901 ACAAAACGANGACATTACATTCAAAACTCTTAATCCAAATATGATATGCCAAGGACAAA 3960
QY 3961 ctatgtctaaagcaagaaaaataaagttagttcttcaatgtctgttttatttcaggacctll 4020
Db 3961 CTATGTCTAAGCAAGAAAAATAAATGTTAGTTCTTCAATGCTCTGTTTTTATTTCAGGACCTT 4020
QY 4021 tcagatttctcttgatacacctttttaggttctgattccagtgagtggaagcacactg 4080
Db 4021 TCAGATTTCTTGGATACCTTTTGTAGGTTCTGATTTCAGTGCAGTGGGAAGACACACTG 4080
QY 4081 actctgacttcaaatagttattacttgcgaatcacattacaacccaaactatcataatca 4140
Db 4081 ACTCTGACTTCAAAATAGTATTACTTGGCAATACATTTAAACCAACAACTATCATATAATCA 4140
QY 4141 caaatgtatcagcgttaattactgtgtcctacctttgttatcaataaagaatactaaagaag 4200
Db 4141 CAAATGTATACAGCTAAATTACTGTGCTCTACCTTTGTATCAATTAAGAAAAATCTTAACAAAG 4200
QY 4201 ttccttgcttaaaaaaataaaaaa 4229
Db 4201 TTCTTGTCTTAAAAAATAAAAAA 4229

RESULT 2
RNO320522
LOCUS

RNO320522 4256 bp mRNA linear ROD 04-SEP-2001

DEFINITION Rattus norvegicus mRNA for complement inhibitory factor H (FH gene).

ACCESSION AJ320522

VERSION AJ320522.1 GI:15485712

KEYWORDS complement inhibitory factor H, FH gene.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 4256)

AUTHORS Demberg,T., Goetze,O. and Schlaef,G.

TITLE Rat complement factor H: molecular cloning, sequencing and expression in tissues and isolated cells

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 4256)

AUTHORS Demberg,T.

TITLE Direct Submission

JOURNAL Submitted (07-AUG-2001) Demberg T., Immunology, Georg August University of Goettingen, Kreuzberggring 57, 37073 Goettingen, GERMANY

FEATURES

source location/Qualifiers

1..4256

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/sex="male"

/tissue_type="liver"

/clone_lib="Lambda Triplex"

/dev_stage="ages 10 - 12 weeks"

/tissue_lib="CLONTECH rat liver 5'stretch plus cDNA library"

24..4209

/gene="FH"

24..3734

/gene="FH"

/codon_start=1

/product="complement inhibitory factor H"

/protein_id="CAC57513.1"

/db_xref="GI:15485713"

/translation="MRLSARIIMLMTVCVVAEDCKGPPPRENSEILLSGSWSEOLYSE
GTOATYKCRPGYRTLGTIVKKNCEWVPSNRICRKPCHGPDTPFGSFLRVLGS
EFEGKAVVYTCDEGYQLLGEIDYREGDAGWNTNDIPICEVVKCLPVTLENGRIVSC
AAEPDQVYFCQVYFECNSGFKIEGOKEMHCSEGLWSNEKPCQVELSCLPVRVNG
DIGYLKPVYKNERFQYCKGQFVYKERGDAVTCGSWNPOPCSEEMTCLTPYIPNGI
YTPRIKIRIDDEIRYCKNGLYPATRSPVSKCTITGWIAPRCSIAKPCDPQPKHGR
LYEESRRYPFPVPIGEYSYCDNGFTPTSPYMDYLRCVTNGWEPEVCLRQCIEH
SDIELNGFTSESDYTYALNRKTRYRCKQGVYNTGETSGIITCLQDQWSPPRSCIKS
CDMPVFENAMTKNNNTWKNDKLDYECHEIGYENYKHTKGSITCTYDQWSSTPSCYE
RECSIPLLHODLVVFPREVYKVGDSLFSFCSCHRRVGAULVQCYHFGWSNPFTECG
QVKSQDQPLETNGEIKGTKKVYSHGDVVEYDCKPRFLKGNPKIOCVDGKWTLPI
CVERCTGDLPELHGSVLSVPPYHHGDSVEFTCTETFTMIGHAVVFCISGRWTEL
PQCVATDLEKAKPSTGIDAIHPNKNFNNHFSVSYRCRQKQYEHISICINRWDEL
EPNCRNEKRFPPQPIPAQVITETTVKYLDEKSVLQDQGYLTQCPHEMWCKHGR
WQSLPRCTEIKPSQPPKIEHGSIKSPRESERDLIESSYEHGTTFSYVCDQGRFLS
EENRVTNMGKWSLPCVGIPOGPPPSIDPLIVSHELESYQVEEVYVNCSEBFGID
GPATKCVGGQWSEPPKCIKTDNDLPTFEIKAPTEKKSRSRGQVTRCPPIRPM
DGSDIVTCVNTKIGQVPCNDSCVNPFPNATILTRHKTKYPSGDKVRYDKNKPYE
LFWGSGDVPTKFTWEPKPKDSTGCGPPPDINDGDITSLPLVPYAPLSSVBYQCO
NYLLKGNKIVTCRNKWSQPTCLHACVIPEDIMEKHNIVLWRRENAKIYVSQSGENI
EFMKPKGYRFRGSPPRPTKCTIEGHINPTCV"

polyA_signal 4194..4199

polyA_site /gene="FH"

4209

/gene="FH"

BASE COUNT 1423 a 788 c 892 g 1153 t

ORIGIN

Query Match 95.28; Score 4025.4; DB 10; Length 4256;

Best Local Similarity 98.33; Pred. No. 0;

Matches 4171; Conservative 0; Mismatches 46; Indels 28; Gaps 9;

QY 10 ctgctccagatagatcccaagacatgagactctcaagaagaattatttggcttatatt 69

DB 1 CTGCTCCAGATAGATCCCAAGACATGAGACTGTCCACAGAATATTTCGCTTATATTAT 60

QY 70 ggactgtttgttagcagaagatltgtaaggctcctctcccaagagaaaaattcagaatttc 129

DB 61 GCACGTGTTTGTCTAGCAGAAGATTGTAAAGGTCTCTCTCAAGAGAAAAATTCAGAAATTC 120

QY 130 tctcagggttcgtggctggaactatctcagaaggcactcaggcaactcacaatgac 189

DB 121 TCTCAGGTTCTGGTCTGAACAACATATATTTCAGAGGGCACTCAGGCAACCTCAAAATGCC 180

QY 190 gccctggataccgaacacttggactattgtaaaagtatgcaagaatgagaaatgagta 249

DB 181 GCCCTGGATACCGAACACTTGGTACTATTGTAAAAGTATGCAAGAAATGAGAAATGGGTAC 240

QY 250 ctcttaaccatcaaggatgctcggaagaaggccatgtgggcatcccgagagacacacct 309

DB 241 CTCTAACCCATCAAGGATATGTCGAAAAGGCCATGTGGCATCCCGAGACACACCCCT 300

QY 310 ttgggtccttttaggctggcagttggaatctgaatttgaatttgggtcaaaaggtgtttata 369

DB 301 TTGGGTCTCTTTAGGCTGGCAGTTGGATCTGAATTTGAATTTGGTCCAAAAGTTGTTTATA 360

QY 370 catgtgaaagggtaccacactattagtgaaattgattaccgtgaatgtgagcagatg 429

DB 361 CATGTGATGAAGGGTACCACACTCTTAGGTGAATTCATTACCGTGAATGTGATGCAGATG 420

QY 430 ggtgaccaatgatattccaatgtaagttgtgaagttgtgagtgcttgcagtgacagaactg 489

DB 421 GGTGACCAATGATATTCCAATATGTGAAGTTGTGAAGTGTGTGCAAGTGTGACAGAATCTG 480

QY 490 aqaatggaagaattgtgagtgctcagccgaacacagaccagaataattattttggaacag 549

DB 481 AGAATGGGAAGAAATGTGAGTGTGAGCCGCAACACAGCAGGAATATATTTTGGACAGG 540

QY 550 tggtagcgtttgaatgcaactccgcttcaagattgaagacagagaaagaatgcaactgt 609

DB 541 TGCTACGGTTTGAATGCAACTCCGCTTCAAGATTGAAGCAGAGAAAGAAATCCACTGCT 600

QY 610 cataaaatggcctctggagcaatgaaaggccaagtgtgtgaaatctcttccctgaccac 669

DB 601 CAGAAAAATGGCCTCTGGAGCAATGAAAGCCACAGTGTGTGAAAAATTTCTTCCCTGCCAC 660

QY 670 cagcagttgaaaatggagatgatatctcgaaccagtttacaagagagaatgaaagat 729

DB 661 CACGAGTTGAAAATGGAGATGATATATCTCAAAACCAGTTTACAAAGGAGAAATGAAGAT 720

QY 730 tccaataaaatgtaagcaaggttttgtgtacaaagaagaaggagatgctgctgcacgg 789

DB 721 TCCAATATAAATGTAAGCAAGGTTTGTGTACAAAAGAAAGAGGCGCATGCTGTCTGCACGG 780

QY 790 gttctggatgaactcctcagcttctgtgaagaatgacatggttgactccatatattc 849

DB 781 GTTCTGGATGAAATCCTCAGCCTTCTGTGAAGAAATGACATGTTTGACTCCATATATTC 840

QY 850 caaatggttatctacacactcacagatgtaaacacagaattgatgataatcagatatg 909

DB 841 CAAATGGTATCTACACACTCACAGATTAAACACAGAAATGATGATGATGAATCAGATATG 900

QY 910 aatgtaaaatggtctctcctcgaaccccgatcacctgtttcacaagtgtacaattactg 969

DB 901 AATGTAAAAATGGGTTATATCTGCAACCCGATCACCTTGTTCAAAGTGTACAATTAATG 960

QY 970 gctggtcctcctcagaatgtagctgaaaccttgatgttttcccaaatcacaacatg 1029

DB 961 GCTGGATCCCTGCTCCCAAGATGTAGCTTGAACCTTGTGATTTTCCCAATTTCAACATG 1020

QY 1030 gacgtctgtattagaagaagccgagacaccttacttccccagctaccctataggaagsgt 1089

DB 1021 GACGCTGTATTATGAAGAAAGCCGAGACCCCTACTTCCCAGTACCTATAGGAAGGAGT 1080

Db 3238 AATATCCATCTGGTGACAAAGTACGTTATGACTGTAATAAACCTTTTCAATATTATTTGGG 3297
Qy 3303 gaagtgaagtgaatgtgccc--aaacaggatttgacagaaacaccacccgaaatgcaaaagatt 3360
Db 3298 GAAGTGAAGCTGATGTGCCAAAAGCGGATTTTGGACAGAAACACCCAGAAATGCAAGACTT 3357
Qy 3361 caacaggagaatgtgggctcctccacctatttgacaaatggagagacatcacctcttctgcat 3420
Db 3358 CAACAGGAAATGTGGGCTCTCCACCTATTGACAAATGGAGACATCACCTCCCTGTGCAT 3417
Qy 3421 taccagatgatgcaccattatcatcagttgtaataatcaatgcccagaactatcatclactta 3480
Db 3418 TACCAGTATATGCCACATTTATCATAGTTGAATATCAATGGCAGAACTATTATCTTACTTTA 3477
Qy 3481 agggaaataacatgataacatgtagaataatggaatggctcagccaccacacccctgcttac 3540
Db 3478 AGGGAATTAAGATAGTACATGTAGAAATGGAAGTGGTCTCAGGCCACCAACCTGCTTAC 3537
Qy 3541 atgcattgtgataccagaagatatattaaggaaaacataataatagttctcagatggaaggg 3600
Db 3538 ATGCATGTGTATACAGAAAGATATTATGGAAGAAACATAATATGTTCTCAGATGGAGGG 3597
Qy 3601 aaaaatgcaagaatttatcccaatcaggaggagaatattgaattcattgtgtaaacctggat 3660
Db 3598 AAAATGCAAGGATTTATCCCAATCAGGCGAGATATTGAATTCATGTTGAACCTGGAT 3657
Qy 3661 atagaaaatcagaggatcacctccgcttcgtacaaaatgagcattgagggcaccatcaatt 3720
Db 3658 ATAGAAAATTCAGAGGATCACCTCGCTTCCGTACAAAGTGCATTCAGGCTCACATCAAT 3717
Qy 3721 atcccactgtgtataaaatcgcatacaaatatttagtaaaaccltatgagtg----- 3772
Db 3718 ATCCCACTTGTGTATAAATCGCTATACAAATATTATAGTAAACCTTATGATGAACCTTTG 3777
Qy 3773 ---agaaatgacatgatatattactaatcacagtttgaatttaccattt-aaatattgtta 3828
Db 3778 TTTAGAAATGCCACATGTATATTACTAATACAGTTTGAATTTACATTTGAATATTCTTTA 3837
Qy 3829 gctcatttctctctaataagatatataaaactttttttatatgtgtggttaac-agaactttt 3887
Db 3838 GCTCATTTCTCTTAATAAGTATATAAACTTTTATATGCTGTTTAATCAAGTAACCTTT 3897
Qy 3888 acagactgttgccaaagaacacattatcatcctaaactcctaacc-aaatatgata 3946
Db 3898 ACAGACTGTTGCCAACAAAGCAAGAACATTAATCAAACTCTTAATCCAAATATGATA 3957
Qy 3947 tgccaaagcaaaactatgctcctgaagcaagaaataaatgtagttctctcattgtctgttt 4006
Db 3958 TGCCAAAGCAAACTATGCTTAAGCAAGAAATAAATGTTAGTTCTTCAATGCTCTTT 4017
Qy 4007 ttattcaggaccttcagatttcttggtacaccttttgt---taggtctgtatcacagtg 4064
Db 4018 TTATTCAGGACTTTTCAGATTTTCTTGGATACCTTTTGTGTAGGTTCTGATTCACAGTG 4077
Qy 4065 agtggagaacacactgactgactgactcacaattagttacttacttgcacatacaacaca 4124
Db 4078 AGTGGAAAGACACACTGACTCTGACTTCCAAATAGTATTACTTGCCAATACATACACCA 4137
Qy 4125 aactatcataatcacaaatgtatcacagctaatctactgtgtctactcttctgtatcaata 4184
Db 4138 AACTATCATATATACAAATGTATACAGCTAATTAATGCTGCTCTACCTTTGTATCAATA 4197
Qy 4185 aagaaatctaagaagttctgtcttaaaaaaataaaaaa 4229
Db 4198 AAGAAATCTAAGAAAAAATAAAAAAAAAAAAAAAAAAAAAA 4242

RESULT 3
MUSH
LOCUS 4300 bp mRNA linear ROD 22-APR-1994
DEFINITION Mouse CFh locus, complement protein H gene, complete cds, clones
MH(4-8).
ACCESSION M12660

M12660.1 GI:193724
complement protein H; protein H; serum glycoprotein.
Mus musculus (strain C57/B10.WR) male liver cDNA to mRNA; Mus
musculus (strain C57/B10.WR) male liver DNA; and Mus musculus
(strain C57/B10.WR) male liver cDNA to mRNA.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4300; 1 to 3425; 3474 to 4300)
Kristensen, T. and Tack, B.F.
Murine protein H is comprised of 20 repeating units, 61 amino acids
in length
Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967 (1986)
86233353
Draft entry and clean copy sequence for [1]. [Unpublished (1986)]
Scripps Clinic and Res Found, La Jolla, CA 92037] kindly provided
by
T. Kristensen, 28-JUL-1986.
FEATURES
Location/Qualifiers
1. 4300
/organism="Mus musculus"
/strain="C57/B10.WR"
/db_xref="taxon:10090"
/sex="male"
/tissue_type="liver"
<1..3425
/note="H protein"
/number=1
101..154
/note="H protein signal peptide"
Join(101..3425,3474..3853)
/note="precursor"
/codon_start=1
/product="complement protein H"
/protein_id="AAA37759.1"
/db_xref="GI:387181"
/translation="MRLSARIILWLMTVCAEDCKPPPRENSEILSGSWSEQLYPE
QFEAGKVVYTCITIVKCKNGKWAASNPRIKCKPKCGHGDTPFGSFLAVGS
GFEAGKVVYTCITIVKCKNGKWAASNPRIKCKPKCGHGDTPFGSFLAVGS
AAETQEVYFGQVRFECNSGFKIEGHKIEHCSNGLWSSOPCEKRSPPYILNC1
DGINKAPVIKENERHYKCKGKGVKPERGDVACTGSGWSSOPCEKRSPPYILNC1
YTPHRIHRSDDIEIHYECNIGFYPTGTSVSKCTPTGWIIPVRCFLKCEFPQFKYGR
LYVESLRPNFVPSYGNKYSYKCDNGFSPSGYSWDLKCTACGMEPEVPCVKCKVFH
VYVNGDSAYWEKVVQGSQKCYNGYSLONGQDPTMTCTENGWSPPCIRIKTCSA
SDIHIDNGFLSESSSIALNRETYSRCKQGYNTGEGISITCLONGWSPPCIRIK
CMPVFENSTIKNTWFLNDKLDYECVGFENYKHTKSGITCYTGWSPGPTCKG
RCSVPTLDRKLVYSPRKYRVDLLEFSCSHGRVDPVOCYHFGWSPGPTCKG
QVASCAPPLEILNGEINGAKKSVYSHGEVYKDCRPFLLKGNKIQCVGNMTLPV
CIEERTCGDIPLEHGSACKSVPPYHHGDSVEFICEENFTMHGHSVSCISGKWTUL
PKCVATDQLEKCRVLKSTGTEIAIKPLTEFHNSTMDYKCRDKQOEVSICINGKWD
EPNCTSKTSCPPPPQIPNTQVIETTVYLDGKLSVLCODNYLTQDEEYKQDGRMQ
SLPRGLEKIPCSOPPTIEHGSINLPRSEERDSIESSSHEHGTTFSYVCDGDFRIPE
ENRITCYMGKNSPTPRCVGLPCGPPPSIPLGTVSLELESYOHGEEVYHCTSGFDIG
PAPFICEGKMSDPPKCIKTDQVLPVKNAILRGKSKSVYRTEQVTFRCQSPYOMN
GSDTVTCVNSRWITKPCRDSTGCGPPPIDNGDITSLSPVTEPLSSVEYQCCQKY
FGQVEVNCENGITWPKPCRDSTGCGPPPIDNGDITSLSPVTEPLSSVEYQCCQKY
YLLGKRTICTNGKSEPTCLHACVIPENIMESHNIILKWRHTEKIYSHSGEDIEF
CKYGYKARDSPFRTKCTINGTINPTCV"
Join(155..3425,3474..3850)
/note="H protein"
/product="complement protein H"
3426..3473
/note="H cds intron"
3474..4300
/number=2
BASE COUNT 1401 a 814 c 902 g 1183 t
ORIGIN 554 bp upstream of XhoI site; chromosome 1.

Query Match 75.08; Score 3172.4; DB 10; Length 4300;
Best Local Similarity 85.68; Pred. No. 0;
Matches 3622; Conservative 0; Mismatches 551; Indels 0; Gaps 6;

Db 2229 ATGGAGATTCACTGGAGTTCATTTCTGAGAGAAACTTCACAAATGATTTGACATGGGTGAG 2288
Qy 2221 tttcttgcatagtggaagtgagccgagcttcctcaatgtgttgcaacagatcaactgg 2280
Db 2289 TTTCTTGCAATAGTGGAAATGGACCCAGCTTCCTTAATGTGTTCACACAGACCAACTGG 2348
Qy 2281 agaaagtlaaagcccccgaagtcacacggtacagatagatgcgaattccataaagaatgaat 2340
Db 2349 AGAAGTGTAGAGTGTGAGTCAACTGGCATAGAGCAATAAACCCAAAATTTGACTGAAT 2408
Qy 2341 Ltaacataaacttagtgagttacagatgtagacaaagcagagatgaacacatcaaa 2400
Db 2409 TTACGCCATAACTCCACCATTGACAAATGTAGACAGCAAGCAGGATACGAACCCCTCAA 2468
Qy 2401 tctgcatcaatggaagatggatctcctgaacccaaactgtacaaagcaaaaagattctgcccctc 2460
Db 2469 TCTGTATCAATGGAAATGGGATCTGAACCAAACTGTACAAAGCAAAACATCTCGCCCTC 2528
Qy 2461 ctccccccagatccaaatgcccaagtgaattgaaccacacgctgaatcttgatgag 2520
Db 2529 CTCCACCCGAGATTCCAAATACCCAAAGTGAATGAACACCACCGTGAATTAATCTGGATGGAG 2588
Qy 2521 aaaaagtattctgttcttggcaagatggttacctaactcagggccagagaagaatggat 2580
Db 2589 AAAAATTATCTGTTCTTGGCAAGACAATTAACCTAACCTCAGGACTCAGAAGAAATGGTGT 2648
Qy 2581 gtaaacatggaagtgagtgagtcgtttaccacgctgcacggaaaaaatttccatgttcccagc 2640
Db 2649 GCAAAGATGGAAGTGGCAGTCAATTACCTCGCTGCATTTCAAAAAATTCATGTTCCTCCAGC 2708
Qy 2641 cccctlaaaatgaacatgagatgatttaagtgcgcccaggtctctcagaagaagagagaat 2700
Db 2709 CCCCTACAAATAGAACATGGATCTATTAATTTACCCAGATCTTTCAGAGAAAGAGAGAT 2768
Qy 2701 taattgagtcgagcagttatgaacacggaactacattcagctattgtcgtgagatgagat 2760
Db 2769 CCATTTGACTCCAGCAGTCATGAACATGCAACTACATTCAGCTATGCTCTGTGATGATGTT 2828
Qy 2761 tcaagatctcgaagaaataatgggttaacctgcaacatgggaataatggagctctctgctc 2820
Db 2829 TCAGGATACCTGAAGAAATAGGATAACCTGCTACATGGGAAATGGAGCACTCCACCTC 2888
Qy 2821 gttgttggaataactctgttgagccccacacttcaattccctcttggtattglltctctatg 2880
Db 2889 CTTGTTGGACTTCCTTGTGGACCTCCACCTTCAATTTCTCTTGGTACTGTTTCTCTGTG 2948
Qy 2881 aactagaagttaccaatagagaggggttacalacaattgttctgaaggctcttcttga 2940
Db 2949 AGCTAGACAGTTACCAACATGGGAAGAGGTTACATACCATTTGTTCTACAGCTTTGGAA 3008
Qy 2941 ttgatgacagcatttattaaatgtgttaggagacagtggtctggaacttcccaaatgca 3000
Db 3009 TTGATGGGACCAAGCATTTATATGCGAAGGAGGAAAGTGGTCTGACCCACCAAAATGCA 3068
Qy 3001 taaaactgattgtgacaacttgcacacatttgaaattgcgaacccgacagaaagaa 3060
Db 3069 TAAAAACGGATTGTGACCTTTTACCCACAGTTAAAAATGCCATAATAACAGGAAGAGCA 3128
Qy 3061 aaaaatcacaggttcaggaagaacaagtgcatttcagatgtccacctcgtatcgaaatgg 3120
Db 3129 AAAAATCATATAGGACAGAGGAACAAGTGCATTCAGATTCAGATGTCAATCTCCTTATCAATGA 3188
Qy 3121 atgacctgcacattgtcacatgtgtaatacgaagtgagattggacagccggtatgcgaag 3180
Db 3189 ATGGCTCAGACACTGTGCATGTGTAAATAGTCGGTGGATTTGGACAGCCGATATGCCAAG 3248
Qy 3181 ataattcctgtgtgaatccaccacatgtgcgaatgctactataactaaagggcacaga 3240
Db 3249 ATAATTCCTGTGTGATCCACCATGTCGCAAAATGCTACTATAGTACAGAGGACCAAGA 3308
Qy 3241 ctaaaatccatctcgttgacaagaatgaattatgactgtaataaaccttltgaattattg 3300
Db 3309 ATAAATATCTACATGGGTGACAGAGTACGTTATGAATGTATTAATAAACCITTTGGAACATTTG 3368

C/ 3301 gggaagtgaagtgtgtgtccaaaacgggatttggacagaacaccggaatg----- 3352
Db 3369 GCCAAGTGAAGTGTGTCTGAAATGGATATGGACAGAAACCAAAAGTCCCGAGGTC 3428
Qy 3353 -----caaaattcaacagggaat 3372
Db 3429 TGTAAATTCGACTTGAAGTCTCAAAACCTTCAAAATGTTTCTCTTAGACTCAACAGGGAAT 3488
Qy 3373 ggggacctctccaccctatlgacaalggagacatcacctcctgttcattaccagtatatg 3432
Db 3489 GTGGCCCTCTCCACCTATTGACAAATGGAGACATCACCTCCCTTGTCATTACCAGTATATG 3548
Qy 3433 caccattatcatcagltgaatalcaatgccagaactattatctacttaagggaataaaga 3492
Db 3549 AACCATTTATCATCACTTGAATATCAATGCCAGAGTATTATCTCCTTAAGGGAAAGACA 3608
Qy 3493 tagtaacatgtagaatggaagtggtctcagccaccacacctgcttacaalgcattgtga 3552
Db 3609 CAATAACATGTACAAATGGAAAGTGTCTGAGCCACCAACATGCTTACATGCTGTGTAA 3668
Qy 3553 taccagaagatattatgaaaaacataatagttctcagatggagggaaaaatgcaaga 3612
Db 3669 TACCAGAAAACATTTAGGAATCACACAATATTAATTTCTCAATGGAGACACACTGAAAGA 3728
Qy 3613 ttattcccaatcagggagaattgaattcatgtgtaaaccttgatataagaataattca 3672
Db 3729 TTTATTCCTCATTACGGGGAGGATATTGAATTTGGATGTAAATATGATATTTATAAGCAA 3788
Qy 3673 gaggatccctcgttctgtacaaagtgcatgaggggtcaacatcaatcccaacttly 3732
Db 3789 GAGATTACCCGGCATTTGCTACAAAGTGCATTAATGGCACCATCAATTTATCCCACTTGTG 3848
Qy 3733 tataaaatcgctatatcaattattagaaccttattggaatggaagaaatgcacatgatatta 3792
Db 3849 TATAAAATCAATAATACATTTATTAGTTGATTTTATTGTTTGAAGGACACATGTCATGTGA 3908
Qy 3793 ctatcacagtttgaatttcaatttaaatattgttttagctcatttctcttaataagtatat 3852
Db 3909 CTAATATACTTTTCAATTTGCATTTGAAGTATTGTTTTAACTCATGTCTCTCTCATAAATATA 3968
Qy 3853 aaacttttttataatggttggttaatacagtaactttcacagactgttgcacaaagaaga 3912
Db 3969 ACATTTTGTGTATATGGTGCATTAACCTTTGAACTTTAAAAAATTTGCCAAAATGCAAAAG 4028
Qy 3913 cattacattcaaaactccttaataccaaatagatgttccaaagcaaaactatgltctaaqc 3972
Db 4029 CAGTAATTCAAAACCTCCCTAACTTAAATATGATATGATGTCCAGACACAACTATTTC-AAATC 4087
Qy 3973 aaaaaataaattgtagttcttcaatgtctgtttttattcaggacctt-tcagattttct 4031
Db 4088 AAGAAAGTAGATGAAGTCTTCAACATCTGTTTCTATTTCAGAACTTTCTCAGATTTTCC 4147
Qy 4032 tggataacctt- - -tggttagttctgattccacagtgagtggaagacacactgactctgac 4088
Db 4148 TGGATACCTTTTGATGTAAAGTCTGATTTACAGTGGATAAAGGATATATTGCAGTATTTC 4207
Qy 4089 ttcaaatagttacttacttgc- - -aatacatttaacaaacaaactatcata - - -atatcaca 4142
Db 4208 TTCAAAATTAATGATTTTCCCAAGCATGTACACACCAAACTATCATATATTATATGACT 4267
Qy 4143 aatgtatcacagctaattactgtglectacctt 4174
Db 4268 AATGCATACAAATTAATTTACTATATAAATACTTT 4299

RESULT 4

A91885 LOCUS A91885 2715 bp DNA linear PAT 22-JAN-2000
DEFINITION Sequence 3 from Patent WO9823638.
ACCESSION A91885
VERSION A91885.1 GI:6740762
KEYWORDS

SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 2715)
AUTHORS	Schwaible,W. and Sim,R.B.
TITLE	COMPLEMENT INHIBITOR
JOURNAL	PATENT: WO 9823638-A 3 04-JUN-1998;
FEATURES	SCHWAEBLE,WILHELM (GB); UNIV LEICESTER (GB) Location/Qualifiers 1..2715
source	/organism="unidentified" /db_xref="taxon:32644"
BASE COUNT	863 a 518 C 618 g 716 t
ORIGIN	
Query Match	55.8%; Score 2360; DB 6; Length 2715;
Best Local Similarity	88.7%; Pred. No. 0;
Matches 2715; Conservative	0; Mismatches Indels Gaps 1;
QY 1	tggagtcacactgctcccagatagataccaaagacatgagactgtcagaagaattatttggc 60
DB 1	TCGAGCTCAACTGCTCCAGATAGATCCAGACATGAGACTGTGCGAACAATATTATTGGC 60
QY 61	ttatatgatgaactgttgtagcagaagattttaaggctcctctccaagagaaaaatt 120
DB 61	TTATATTATFGGAAGTGTGTGTAGCAGAAGATTTGTAAGGTCCTCTCCAAGAGAAAAAT 120
QY 121	cagaaattctctcaagttcggtgtgaaccaactatattcagaaggcactcaygcaacct 180
DB 121	CAGAAATTCCTCAGGTTCTGAGTCTGAACAACATAATTACAGAGGCACCTCAGGCCAACCT 180
QY 181	acaaatgcgcccttgataccgaacaacttgttaactattgtaaaagtatgcaayaatygg 240
DB 181	ACAAATGCGGCCCTGGATACCAGAACACTTGGTPACTATTGTAAAAGTATGCAAGAATGGAG 240
QY 241	aatgggttaccttctaaccatcaaggatattgcggaaggccatgtgggcactccccnag 300
DB 241	AATCGGTACTTCTAACCCATCAAGGATATGTCGGAAGAGGCCATGTGGCATCCCGGAG 300
QY 301	acacacccttgggttcctttaggctggcagttggactgaatttggatggcacaagg 360
DB 301	ACACACCCTTTGGGTCTCTTAGCTGGCAGTTGGATCTGAATTTCAAATTTGGTGCAAGG 360
QY 361	tgttttatcatatgatatgaagggttaceaacattataggatgaattgatlacctgtaattgy 420
DB 361	TTGTTTTATACATGTGATGAAGGGTACCACACTATTATAGGTGAAATTCATTACCCTGAATGTG 420
QY 421	atgcagatggglggaccaatgatattccaatatgtgaaqtlgtgaagtgcttlyccaqlga 480
DB 421	ATGCAGATGGGTGGACCAANTGATATCCAATATGTGAAAGTTGTGAAGTGTCTTGCCAGTGA 480
QY 481	cagaactgggaatyggaagaattgtgagtggtgcagccggaaccagaccaggaattatt 540
DB 481	CAGAACTGGCAATGGGAAGAATTGTGAGTGTGTGACGCGCAACCAAGCACGAGGAATATTATT 540
QY 541	ttgacacagggtgacgcttttgaatgcaactccgctccaagattgaaaggacagaagaaa 600
DB 541	TTGACACAGGTGGTACGCTTTTGAATGCAACTCCGCTTCCAAGATTGAAGGACAGAAAGAAA 600
QY 601	tgcactgctcataaaatggcctctggagcaaatgaaaagccacagtgtgtgaaaatttctt 660
DB 601	TGCACCTGCTATAAAATGGCCCTCTGGAGCAATGAAAAGCCACAGTGTGTG ----- 650
QY 661	gcctgccaccacgagttgaaaaatggagatggttatatatctgaaaccagtttacaaggaga 720
DB 651	----- 650
QY 721	atgaaagattccaataataaatgtaaagcaagggttttgtgtacaaagaaagagggatgctg 780
DB 651	----- 650
QY 781	tctgcacgggttctggatggaatcctcagccttctgtgaagaaatgacattgtltgactc 840

Db 1516 GCTACCACCTTTGGATGTCCTCCCTAATTTCCCAACGTGTGAAGGCCAAGTAAATCATGTG 1575
Qy 1921 accaacctcttgaaatcccgaatgggaaataaaggagcaaaaaaagttgaaatcacgccc 1980
Db 1576 ACCAACCTCTTGAATCCGGAATCGGCAATAAAGGAGCAAAAAAACTTGAATACAGCC 1635
Qy 1981 atggtgacgtggtggaatatgatgtgcaaacctagattcttactgaaaggacccaataaaa 2040
Db 1636 ATGCTGACCTGCTGGAATATGATTCGCAACCTAGATTCTACTCAAGGGACCCCAATAAAA 1695
Qy 2041 tccagtgattgacggaagtggacaagttggcgatatcgatcggttgagatgagagaacat 2100
Db 1696 TCCAGTGTGTGACGGGAAGTGGACAAGTTGGCGGATATGCGTGTGAGTATGAGAGACAT 1755
Qy 2101 gtgagagaccttcctggaacttgagcatggctctgtcaagttatctgtcctccctaccatc 2160
Db 1756 GTGAGAGACCTTCCCTGAACCTTGAGCATGGCTCTGTCTCAAGTTATCTGTCCCTACCATC 1815
Qy 2161 atgagagattcaagtgttcaacttgatcacagaaccttcacaaatgattggacatgcaglaq 2220
Db 1816 ATGAGAGATTCACTGGAGTTTCACTTGTACAGAAACCTTCACAATGATTGGACATCCAGTAG 1875
Qy 2221 tttcttcattagtggaagggtggaccagcttctcctcaatgtgttgcaacagatcaactgg 2280
Db 1876 TTTTCTGCATTAGTGAAGTGGACCGAGCTTCTCTCAATGTTGTGCAACAGATCAACTGG 1935
Qy 2281 agaagtglaagcccccgaagtcgaactggcatagatgcaattatcctccaaaataagaatgaat 2340
Db 1936 AGAAGTGTAAAGCCCGCAAGTCAACTGGCATAGATGCAATTCATCCAAATAAAGAAATGAAT 1995
Qy 2341 ttaatcataacttttagtgaagttacagatgtagaacaagcaggagatgagaaatcaca 2400
Db 1996 TTAATCATATACTTTAGTGTGAGTTACAGATGTAGACAAAACGAGGATATGAACATTCAA 2055
Qy 2401 tctgcatcaatggaagtggaatcctgaaacaaactgtacaagcaaaagattctgcctc 2460
Db 2056 TCTGCATCAATGGAAGATGGATGCTTCAACCAAACTGTACAAAGCAAAAGATTCTGCCCTC 2115
Qy 2461 ctccccgagagattccaaatgcccgaatgattgaaacacacgtagaaatacttggaatgag 2520
Db 2116 CTCCCCGCGAGATTCCAAATGCCCAAGTGATTGAACACACCGTGAAATACTTGGATGCAG 2175
Qy 2521 aaaaaglatctgttcttgcgaagatggttacctaactcagggcccgagaagaatgagat 2580
Db 2176 AAAAAAGTATCTGTTCTTGTGCAAGATGGTTACCTTAACCTAGGCGCCCAAGAAAGATGTTG 2235
Qy 2581 gtaacacatggaagtggaatcggttaccacgctgcacgaggaataatccatgcttcagc 2640
Db 2236 GTAAACATGGAAGTGGCAGTCGTTACCCAGCTGCACGGAAAAAATTCATGTTCTCCAGC 2295
Qy 2641 cccctaataatgaacatggaatctattaagtcgacccagttcctcagaagagagagagat 2700
Db 2296 CCCCCTAAATTTGAACATGGATCTATTAAAGTCGCCCCAGTCTCTCAGAGAGAGAGAGATT 2355
Qy 2701 taattgagtcagcagttatgaacacggaactacattcagctattctgctgagagagagat 2760
Db 2356 TAATTGAGTCCAGCAGTTATGAACACGGAATACATTCACTATTCTGTAGAGATGCAT 2415
Qy 2761 tcaagatatctgaagaaaaatagggttaacctgcacaatggaataatggagctctctcctc 2820
Db 2416 TCAAGATATCTGAAGAAAAATAGGGTAACCTGCAACATGGGAAAAATGAGCTCTCTGCCCTC 2475
Qy 2821 gttgtgtggaataacttggagcccccccaacttcaattcctcttggtattgtttctcatg 2880
Db 2476 GTTGTGTTGGAATACCTTGTGGACCCCCCTTCAATTCCTCTTGTGATTGTTTCTCATG 2535
Qy 2881 aactagaaagtaccataatggagagaggttatcacataacttctctggaagccttttggaa 2940
Db 2536 AACTAGAAAGTTACCAATATGGAGAGAGGTTACATACAAATGTTCTGAAGGCTTTGGAA 2995
Qy 2941 ttgatgaccagcattattataatgtgtaggagagcagtggttgaaacctccccaaatgca 3000
Db 2596 TTGATGGACGAGCATTTATTAATGTGTAGGAGGAGAGTGGTCTGAACCTTCCCAATGCA 2655

Qy 3001 taaaaactgattgtgacaacttcccacatttgaattgaaattgcaaacccacacagaaagaaa 3060
Db 2656 TAAAAACTGATTGTGACAACTTGCCACATTTCGCAAACTTGAATTTGCCAAACCCACACAAAAA 2715
RESULT 5
HSH LOCUS HSH 3926 bp mRNA linear PRI 08-APR-1997
DEFINITION Human mRNA for complement factor H.
ACCESSION Y00716 M32093 X07525
VERSION Y00716.1 GI:31964
KEYWORDS complement factor H; glycoprotein; regulatory protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3926)
Direct Submission
Submitted (25-FEB-1988) Day A.J., Harris,T.J. and Sim,R.B.
of Biochemistry, University of Oxford, South Parks Road, Oxford,
OX1 3QU
2 (bases 1 to 3926)
Ripoche,J., Day,A.J., Harris,T.J. and Sim,R.B.
The complete amino acid sequence of human complement factor H
Biochem. J. 249 (2), 593-602 (1988)
88134059
COMMENT see Immunogenetics 27:211-214(1988) concerning TYR <-> HIS at AA
384
see X07523 for truncated (1.8 kb) mRNA sequence.
FEATURES
Location/Qualifiers
1..3926
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="B-38-1, R2a and PE3"
/tissue_type="liver"
/clone_lib="human liver cDNA in pat153/PvuII/8"
74..127
74..3769
/codon_start=1
/product="factor H"
/protein_id="CAA68704.1"
/db_xref="GI:31965"
/db_xref="SWISS-PROT:P08603"
/translation="MRLLAKIIICLLMWAICVAEDCNELPPRRNTEILTGSWSDOTYPE
GTQAIYKCRPGVRSICNVIMVCRKGWALNPLRCKQKPCGHPGDTPTCTTLTGNN
VFEGYGVKAVYTCNEGYYQLLGEINVRCDTDDGWTNDIPICEVVKCLPVTAPENKIVSS
AMEPDREYHFGQAVRFVCSGYKIEGDEMHCSDDGFNFKPKVCEISCKSPDIVNG
SPIQSUKIYKENERFOYKCNMGVEYSERGDAVCTESGWRPLPSCKEKSCDNPIYINPD
YSLIRIKHRTGDEITYOCNRGEYPATRGNTAKCTGTGTPAPRCTLKPCDYPDKKHGG
LYHENMRPYPVAVGKYYSYCDHEFTFPGSYWDHIHCTODGWSFVAPVCLKKCYFP
YLENGYNQHKRFVQKSIDVACHPGYALPKAOTVTCTMENGWSPTPRCIRVKTCCK
SSIDIEGFLISEOYTYALKKAKYCKLGYVTADGETSGSIRCGKDGWSAOPTCKS
CDIPVFNARTKNDFTWPKLNDTLDECHDGYESNTGSTTGSIVCGYNGWSDLPICTE
RECELPKIDVHLVPRDKKDOYKGEVLKTSCKPGFTIVGNSVOCYHIFGLSPDLPICK
EQVQSGPPPELLNGVYKETEKEYEYSEVYECNPRFLMKGPNKIQCVDGSMWTLIP
VCIVBEESTGDIPELHGWAQLSSPPYYGDSVEFNCSSEFTMIGHRSITC IHGVWTFQ
LPQCAVIDKLKKCKSSNLIILEHLKNKKEFHDNSNIRYRCRCKGWIHTVC INGRWD
PEVNCMAOIQLCPPPPOIPNSHNWTTILNRRDGEKVSILCOENYLIQEGEELTKDQ
KQMSIFLCVEKIPCSQPIEHGTINSSRSSQSEYAHGKLSITCEGGFRISENETT
RYMGSPPQCEGLPCKSPPEISHGVVAHMSDSYQGEVYTKCEGFGIDGPAIAK
CLGKRWSPSCITKTDCLSLPSFENAIPMGKKDYKAGEQVYTCATYTKMDGASNV
TCINSRWTRPTCTSCVNPPTVONAYIVSRMSKYPGSERVYRCRYPYEMFGDEE
VMCLGNMTPEPQCKDCTCKGPPPPIDNGDITSFPLSVYAPASSVEYQCNLYOLEG
NKRTICRNGQSEPPKCLHPCVISRIMENYNIALRWTAQKLKLSYRSGSEVFCVRG
YKLSRSHTLRITTCWDGKLEYPTCAAR"
128..3766
/product="factor H"
994
/note="polymorphism"
/replace="-C"
1277
sig_peptide
CDS
mat_peptide
variation
variation

[illegible]

Db 1966 TCACCTCTGAACTCCTCAATGGGAATGTTAAGGAAAAAACGAAAGAATATGCACA 2025
Qy 1982 tggctgctggtggaataatgattgcaaacctagatttctactgaaggagcccaataaaaa 2041
Db 2026 CACTGAAGTGGTGAATATTATTGCAATCCTACATTTCTAATCAAGGACCTAATAAAT 2085
Qy 2042 ccaagtgttgacggaagtggacaaggttgccgatatatgcgttgagatgagagaacatg 2101
Db 2086 TCAATGTGTGTGAGAGAGTGGACAACTTTACCAGTGTGTATTGTGTGGAGGAGTACCTG 2145
Qy 2102 tggagaccttccctgaacttgagcatggctctgtcaagttatcttgcctccctaccatca 2161
Db 2146 TGGAGATATACCTGAACTTGAACATGGCTGGGCCAGCTTTCTCCCTCTTATTACTPA 2205
Qy 2162 tggagattcaagtggagttccactgttcagaaaaaccttcacaatgattggacatcgatgt 2221
Db 2206 TGGAGATTCACTGGGAATTCAAATTTGCTCAGAAATCATTTACAATGATTGGACACAGATCAAT 2265
Qy 2222 ttctgcattagtggaagtggaccagcttctcctaagtgttggaacagatcaactgga 2281
Db 2266 TAGCTGTATTTCAGGATGTGGACCCAACTTCCCAAGTGTGTGGCAATAGATAAATCTTA 2325
Qy 2282 gaagtgaagcccccgaactggaactggcatagatgcaattctcacaataaagaatgaatt 2341
Db 2326 GAAGTCAAAATCATCAATTTAATTTAATCTAGGAAACATTTTAAAAACAAGAAGCAAT 2385
Qy 2342 taatcataacttttagtgtgagttacagatgtagacaaaagcaggagatgtaaacattcaat 2401
Db 2386 CGATCATAAATTTAAGATAGGTACAGATGTAGAGAAAAGAAAGAGATGGATACACACAGT 2445
Qy 2402 ctgcataatgggaagtggatcctgaaacaaactgtacaag - - - - -caaaagattctcg 2455
Db 2446 CTCATAAATGGAAGATGGATCCAGAGTGAACTGCTCAATGGCAACAATACAATATAG 2505
Qy 2456 cctctctccccgcagattcccaaatgcccaagtgtatgaaacacccgtgaaatcacttga 2515
Db 2506 CCACCTCCACCTCAGATTCCCAATTTCTCACAATATGACAAACCCACACTGAATTTATCGGA 2565
Qy 2516 tggagaaaaagtatctgttctttgccaagtgttaccctaaactcagggcccgcaagaagaa 2575
Db 2566 TGGAGAAAAGTATCTGTCTTTGCCAAGAAAATTTATCTAATTCAGAAAGGAGGAAGAAAT 2625
Qy 2576 ggtgtgaacatgggaagtggaagtcgtttacacgctgcacggaaaaaatlccatlytic 2635
Db 2626 TACATGCAAAAGATGGAAGATGGCAGTCAATACCACCTCTGTGTGAAAAAATTTCCATGTT 2685
Qy 2636 ccagccctcaaatgaacatggtatcttaagtgcgccaggtctccgagagagagag 2695
Db 2686 ACAACCCTCAGATAGAACGGAACCATTAATTCATCCAGGTCTTCACAAGA - - - - - 2739
Qy 2696 agatttaattgagtcagcaggttataacacggaactacattcagctattgctgtagaga 2755
Db 2740 - - - - -AAGTTATGCACATGGGACTAAATTCAGTTTACTTTGTGAGGG 2781
Qy 2756 tggattcaagatatctgaagaaaaatagggtgaacctgcaacatgggaaaaatggagctctct 2815
Db 2782 TGGTTTCAGGATATCTGAAGAAAAATGAAACACATGCTACATGGGAAAATGGAGTCTCC 2841
Qy 2816 gctcgtgtgttggaataaccttggaccgcccaaccttcaattcctctgtgttgattgttc 2875
Db 2842 ACCTCAGTGTGAAGGCCCTTCCCTGTGTAATCTCCACCTGAGATTTCTCATGGTGTGTAGC 2901
Qy 2876 tcataactagaagaattaccaataaggagaggaggtttacatacaaatgttctgaaggctt 2935
Db 2902 TCACATGTCAGACAGTTATCAGTATGGAGAACAGTTAGTACAAATGTTTTCAGAGTTT 2961
Qy 2936 tggaaattgatgacagcaatttataatgtgtlaggagacagtggtctgaaacctcccaa 2995
Db 2962 TGGAAATTGATGGCCCTGCAATTTGCAAAATGCTTAGGAGAAAAATGGTCTCACCCCTCATC 3021
Qy 2996 atgcataaaaaactgattgtgacaacttgcaccaatttgaattggcaaacccagacagaaa 3055
Db 3022 ATGCATAAAAAACAGATTGTCTCAGTTTACCTAGCTTTGAAAATGCCATACCCATGGAGA 3081

Qy 3056 gaaaaaaaatcaccaggttcagagagaacaagtgcacattcagatgtgccacctccglatcg 3115
Db 3082 GAGAAGGATGTGTATAAGCGGGTGAGCAAGTGACTTACACTTGTGCAACATATTACAA 3141
Qy 3116 aatggatggctctgcacattgttcacatgttataacgaagtggattggacagccggtatg 3175
Db 3142 AATGGATGGAGCCAGCAATAATTAACATGCATTAAATAGCAGATGGACGAAGGCCAACATG 3201
Qy 3176 caaagataattctcctgtgtaattccaccacatgtgcceaaatgctactatactacaagaagca 3235
Db 3202 CAGAGACACCTCTGTGTGAATCCGCCACAGTACAAAATGCTTTATATATGTGCGAGACA 3261
Qy 3236 caagactaaatctccatctgggtgacaagaatgcattatgactgtataaaaaaccttltgaatt 3295
Db 3262 GATGAGTAAATATCCATCTGCTGAGAGAGTACGTTTATCAATGTAGGAGCCCTTATGAAT 3321
Qy 3296 atttggggaagtgaagtgtatgccccaaacgggatttggacagaaccccccgaatgcaa 3355
Db 3322 GTTTGGGGATGAAGAAGTGTATGTTTAAATGGAAATGGACGGAACCCACTCAATGCAA 3381
Qy 3356 agattcaacagggaatgtgggctcctccacctattgacaattggagacatcacctcctt 3415
Db 3382 AGATTCTACAGGAATAATGGGCCCTCCACCTATTGACAATGGGACATTTACTTCATT 3441
Qy 3416 gtcattaccagtatatgcaccattatcatcagttgaaatcaatgccagaactattatct 3475
Db 3442 CCGCTTGTGAGTATATGCTCCAGCTTCATCAGTTGAGTACCAATGCCAGAACTTGTATCA 3501
Qy 3476 acttaagggaataaagatagtaacatgtagaagaatgaaagtggctcagccaccaacctg 3535
Db 3502 ACTTGAGGGTAAACAGCGAATAACATGTGAATGGACAATGGTTCAGAACCCACCAAAATG 3561
Qy 3536 cttacatgcattgtgtatccagaagatatattggaacacataataatagttctcagatg 3595
Db 3562 CTTACATCCGTGTGTAATATCCGAGAAATTTATGGAATAATTATAACATACCATTAAGGTG 3621
Qy 3596 gagggaaaaatgcaagaattattccccaaatcagggggagaatatattgaattcactgtaaac 3655
Db 3622 GACAGCCAAACAGAAAGCTTTTATTCGAGAACAGGTGAATCAGTTTGAATTTGTGTAAACG 3681
Qy 3656 tggatatagaataatcagaggatcacctcgtttctgacaaaagtcgatlgagggtcacat 3715
Db 3682 GGGATATGCTCTTTCATCAGTCTCACATTTGCCAACACACATGTTGGGATGGGAAACT 3741
Qy 3716 caattatccacctgtgtgtataaaatgcgtatacaattattagtaaaccttattgtagaga 3775
Db 3742 GGAGTATCCAACCTGTGCAAAAAGATAGAATCAATCAATAAAGTGCACACCTTTATTTCAGA 3801
Qy 3776 aatgcacatgtatatactaataacagtttgaatttcaattttaaatttattgtagctcatl 3835
Db 3802 ACTTTAGTATTAAATCAGTTCTCAATTTTCATTTT - - - - -TTTATGTATTGTTTACTCCTT 3856
Qy 3836 tctctcaataaagata 3851
Db 3857 TTTATTTCATACGTAAA 3872
RESULT 6
MUSCFHRD
LOCUS
DEFINITION Mouse complement factor H-related protein mRNA, complete cds, clone 9C4.
ACCESSION M29009 J05259
VERSION M29009.1 GI:192563
KEYWORDS complement factor.
SOURCE Mouse (strain C57/B10.WR) liver, clone 9C4.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2955)
AUTHORS Vik,D.P., Munoz-Canoves,P., Kozono,H., Martin,L.G., Tack,B.F. and Chaplin,D.D.


```
REFERENCE 1 (bases 1 to 1532)
AUTHORS Schwaebler,W. and Sim,R.B.
TITLE COMPLEMENT INHIBITOR
JOURNAL PATENT: WO 9823638-A 4 04-JUN-1998:
SCHWAEBLER,WILHELM (GB); UNIV LEICESTER (GB)
FEATURES
    Location/Qualifiers
        source
            1..1532
                /organism="unidentified"
                /db_xref="taxon:32644"
BASE COUNT 523 a 275 c 290 g 444 t
ORIGIN
    Query Match 26.5%; Score 1119; DB 6; Length 1532;
    Best Local Similarity 100.0%; Pred. No. 3.6e-239;
    Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3111 tatcgaaatgagctgctgcacatgtcacatgtgtaataacgaagtggattguacagccg 3170
DB 414 TATCGAATGGATGGCTCTGCATGTGTGCATGTGTCAATACGAAGTGGATTGGACAGCCG 473

QY 3171 gtatgaaagaataatctctgtgtaataccaccacatgtgcaaatgctactataca 3230
DB 474 GTATGCAAAAGATAATCTCTGTGTAATCCACCAATGTGCCAATGCTACTATACTAACA 533

QY 3231 aggcacaagactaaataccatctggtgacaaagtcattatgactgaataaaccctttt 3290
DB 534 AGGCACAAGACTAAATATCCATCTGTGTGCACAAAGTAGCTTATGACTGTATAAACCCTTT 593

QY 3291 gaattatttgggaagtgggaagtgtgacaaacgggatttggacagaaacacagaaa 3350
DB 594 GAATTATTTTGGGAAGTGGGAAGTCATGTGCCAAACGGGATTTGGACAGAACCCAGAAA 653

QY 3351 tgcaaaagattcaacagggaagtgtggccctctccacatttgcacaaatggagacatcacc 3410
DB 654 TGCAAAAGATTCAACAGGGAATGTGGCCCTCTCCACCTATTGACATGAGAGACATCACC 713

QY 3411 tccttgcataccagtatatgcaccattatcatcatcagttgaaatcaatgccagaactat 3470
DB 714 TCCTTGTCATTACAGTATATGCACCATTTATCATCATCTTCAATATCAATGCCAGAACTAT 773

QY 3471 talctacttaaggaaataagatagtaacatgtagaataaggaaagtgctcagaccaca 3530
DB 774 TATCTACTTTAAGGAAATAGACATAGTAACATGTAGAAATGGAAGTGTCTCACCACCA 833

QY 3531 acctgttcatatgcattgtgtatcaccagaagatattatggaataacataatagttctc 3590
DB 834 ACCTGCTTACATGCATGTGTATACCAAGAGATATTATGCAAAACATAATATAGTTCTC 893

QY 3591 agatgagggaataatcaaaagatttattcccaatcaggggagagaatatgaaatcatglt 3650
DB 894 AGATGGAGGGAATAATGCAAAAGATTTATTCCTCAATCAGGGGAGAAATATGAAATTCATGTGT 953

QY 3651 aaacctggatatagaaaattcagaggatcacctccgtttcgtacaaagtgcattgaggt 3710
DB 954 AAACCTGGATATAGAAATTCACAGGATCACCTCCGTTTGTACAAAGTCATGAGGGT 1013

QY 3711 cacataattatccactgtgtatataaaatcgctatatacaattattagtaaaccttatgga 3770
DB 1014 CACATCAATTTATCCCACTTGTGTATATAAAATCGCTATACAAATTTATTAGTAACCTTATGGA 1073

QY 3771 tggaaatgcacatgtatattactaataacagtttgaatttacatttaaatattgtttagc 3830
DB 1074 TGAAATGCACATGTATATTACTAATACAGTTTGAATTTACATTTAAATATTGTTTAGC 1133

QY 3831 tcattctctataaagatatataaacttttttttatgtggtggttaacagaaacttaca 3890
DB 1134 TCATTTCTCTAATAAGTATATAAATTTTTTATATGTTGGTTAATCAGTAACCTTATACA 1193

QY 3891 gactgttgcacaaagcagaacattacattccaaactcctaataccaaatgatgttc 3950
DB 1194 GACTGTTGCCAACAAAGCAAGAACATTACATTCAAACTCCTAATCCAAATATGATATGTC 1253
```

```
QY 3951 caaggacaaactatgtctaaagcaagaaaaataatgttagttcttcaatgtctgttttat 4010
DB 1254 CAAGGACAAACTATGCTAAGCAAGAAAAATAATGTTAGTTCTTCAATGCTGTTTTAT 1313

QY 4011 tcaggaccttcagattttcttggatacccttttggtaggttctgattccacagtgagtgga 4070
DB 1314 TCAGGACCTTTTCAGATTTTCTTGGATACCTTTTGTAGTTCTGATTTCACAGTGAGTGGA 1373

QY 4071 agacacactgactgacttcaaatagatttagttactgcaatacattacaacaaactat 4130
DB 1374 AGACACACTGACTCTGACTTCAAAATTAGTATTACTTCCAATACATTAACACCAAACTAT 1433

QY 4131 cataatatcacaatgtatcacagctaatctactgtgtcctacccttggatcaataaagaaa 4190
DB 1434 CATAATATCAAAATGTATACACTAATTACTGTGCTCTACCTTTGTATCAATAAAGAAA 1493

QY 4191 tctaaagaagcttctgcttaaaaaaataaaaaa 4229
DB 1494 TCTAAGAAAGTTCTTGTCTTAAAAAATAAAAAA 1532

RESULT 9
HUMCHF HUMCHF 2132 bp mRNA linear PRI 01-NOV-1994
LOCUS Human complement H factor mRNA, complete cds.
DEFINITION M17517
ACCESSION M17517
VERSION 1.1 GI:180497
KEYWORDS H factor; complement; glycoprotein.
SOURCE Human liver, cDNA to mRNA, clone R2a.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2132)
AUTHORS Day,A.J., Ripoch,J., Lyons,A., McIntosh,B., Harris,T.J. and
Sim,R.B.
TITLE Sequence analysis of a cDNA clone encoding the C-terminal end of
human complement factor H
JOURNAL Biosci. Rep. 7 (3), 201-207 (1987)
MEDLINE 88025472
FEATURES
    Location/Qualifiers
        1..2132
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="1q32"
        gene
            1..2132
            /gene="HF"
            <1..1976
            /gene="HF"
        CDS
            /note="complement H factor"
            /codon_start=2
            /protein_id="AAA52016.1"
            /db_xref="GI:180498"
            /db_xref="GDB:G00-120-041"
            /translation="DFRNVDPKDKQYKVEVLKFSCKPGFTTIVGPNVQVCFHGLSP
            DLPCKEQVQSCGPPPELLGNVKNKTEKEIGHSEVEYICNPRFLMKGNKIQCVDG
            EWTTLPGVIEESTGCDTPELEHGAQLSPPYGDSVEFNCSESTMGHRSITCI
            HGVWQLPQCAVDKDKKSSNLIILEHLKKNKFEHNSNIYRRCRGKGIHFPVC
            INGRWDPEVNCMAQIQLPKPPPIPNSHNMVTLNRYDGEKSVLCQENYLIQEGEE
            ITCKDGMQSIPLCYVEKIPCSOPPOIEHGTINSSRSQESYAGHTKLSVTCGGPRIS
            ENEFTCYMGKWSPPQCEGLCKSPPEISHGVVHMSDSYQGEVYTKCEGEGID
            GPAIAKCLGEKWSHPSCKTCLSLPSENAIPMGEKKDVTYKAGEQVYITCATYTKM
            DGAINVTCINSRWTPRTCDRTSCVNPPTVQNAVIVSRQMSKYPGSRVRYQCRSPYE
            MFGDEVCLNGNMTPEPPQCKDSTGKCPGPPIDNGDITFSPLSVAPASVEYQCCQN
            LYQLGNKRITCRNGQWSEPPKCLHPCVISREIMENYIALRWTAQKQKLSYRTGESVE
            FVCKRGYRLSSRSHTLRTTCMDGKLEYPTCAKR"
BASE COUNT 714 a 385 c 454 g 579 t
ORIGIN 92 bp upstream of Avail site; chromosome 1q32.
```

Query Match 24.0%; Score 1016.4; DB 9; Length 2132;
Best Local Similarity 69.5%; Pred. No. 2.7e-216;
Matches 1455; Conservative 0; Mismatches 601; Indels 38; Gaps 4;


```

VERSION X98697.1 GI:1419423
KEYWORDS complement factor H.
SOURCE cow.
ORGANISM Bos taurus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
          Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 208)
           Soares,C.J., Day,A.J. and Sim,R.B.
           Prediction from sequence comparisons of residues of factor H
           involved in the interaction with complement component C3b
           Biochem. J. 315 (Pt 2), 523-531 (1996)
JOURNAL Biochem. J. 315 (Pt 2), 523-531 (1996)
MEDLINE 96202005
REFERENCE 2 (bases 1 to 208)
           Sim,R.B.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-1996) R.B. Sim, University of Oxford, MRC
          Immunochemistry Unit, Department of Biochemistry, South Parks Road,
          Oxford, OX1 3QU, UK
FEATURES source
          1..208
          /organism="Bos taurus"
          /db_xref="taxon:9913"
          /tissue_type="liver"
          /dev_stage="adult male"
          <1..>208
          /codon_start=1
          /product="CCP modules 3-12, with parts of CCP 2 and 13"
          /protein_id="CAA67257.1"
          /db_xref="GI:1419424"
          /db_xref="SPTREMBL:Q28085"
          /translation="VSGSPHLAEGNFGEYGAKVYYTCDGYQMVGEMFNRECDTNGMTW
          NDPICEVVKCLVPTEPNKGFISDALDEPOEYTYGVQVFECSNGLMDGPKOIHCS
          AGWGSAETPCKEIFCIPVLINGVALPKATYKNERYVCRAAGFEGYGGRGDTVFC
          TKSGMTAPATICIEITDPPRIINGVALPELSKYRGDKITYECKKGDFEFIRGTDAVC
          TRDGWPVPRCAKCPYPVIKHRLYSYRGVEPARVNQOFVYSGDPHHPVPSORSBW
          DLHACAEHSPEECPLRQCIFYNLGNHNHBEEXYLOGETVRVHCYEYEGYSLQNQND
          TWTCESGNSPPRCIRIVKTCSNRIRIENGFLSESTFYPLNKQTEYCKPGTVTDAD
          GKTSGILTLKNGDSQAQCIKSDRPVFEPKARKVSKDGTWFLRNLDLYDCVGDYENRR
          DGRTTGISVCGQGWSDAKARECSIPEDMPLYLNAYPRKETYKVDLVLKFSQGRG
          IMYGADVOCYHFQWSPKLPCTCKVKVKSCALPELPNGRKRIHKHEYAHHVEYVA
          CNPKFLMKSHKTKTCQCVDESTALPVGIEERTCEISDLHDGDKVPSVPPIIHGDSVEF
          SCREATFMIGPRITICISGBWTOPPOCIATDELKRCKGSTLFELL"
BASE COUNT      629 a   382 c   465 g   532 t
ORIGIN
Query Match      23.6%; Score 998.4; DB 4; Length 2008;
Best Local Similarity 70.3%; Pred. No. 2.8e-212;
Matches 1399; Conservative 0; Mismatches 576; Indels 15; Gaps 4;

QY 309 ttggggtcccttagctggcagttggatcgatgaatttgaaattggcgaaaggctgttat 368
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4 TCCTGGTTCTCCTCATCTTCGGGAAGAAATCAGTTTGCAATATGTCGAAAGGTTGTTTTAT 63

QY 369 acatgtgatgaagggtaccacactattaggtgaaattgattaccgtgatgtgatgcagat 428
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
64 ACCTGTGCATCAGGGGTATCAGATGGTGGGTGAGATGAATTCCGTGAATGTGACACAAT 123

QY 429 gggtgaccaaatgatattccaatatgtgaagtgtggaagtgttcgccagtgcagaactg 488
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
124 GGATGGACCATAATATATTCCTCATGTGAAGTTGTTTAAGTCTTTACCAGTGACAGAACA 183

QY 489 gagaaatggaagaaattgagtggtgcagccgaaccagacagaaaataatttttggacag 548
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 GAGATGGAATAATTTTCATGTGATGCCCTGGACCAACCAAGAAATATACCTATGGACAA 243

QY 549 gtggtacgctttgaatgcaactccggtctcaagattgaagcagagaagaatgcactgc 608
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
244 GTGGTACAGTTTGAGTGTAAATTCAGGCTACATGCTGGATGGACCCCAACAATACATGTC 303

QY 609 tcataaaatggcctctggagcaaatgaaagccacagttgtgtgaaaaattcttctgctgcga 668
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


TITLE Identification and sequence analysis of 4 complement factor
H-related transcripts in mouse liver
J. Biol. Chem. 265, 3193-3201 (1990)
MEDLINE 90153969
COMMENT Draft entry and computer-readable sequence for [J. Biol. Chem.
(1990) In press] kindly submitted
by D.P.Vik, 13-OCT-1989.
FEATURES
Source Location/Qualifiers
1..2356
/organism="Mus musculus"
/db_xref="taxon:10090"
-74..875
/note="complement factor H-related protein"
Join(74..875,1028..1204,1365..1744)
/partial
/note="complement factor H-related protein"
/codon_start=1
/protein_id="AA37415.1"
/db_xref="GI:387128"
/translation="MGFCNMLLSNILLTAMLSSTAKGEEKTCSPPVILNGIYTPHRII
HRSDDEIRYECNYGYPVTGSVSKTPTGMPVPRCTLKPEFPQFYGRLYYEESL
RPFVPSIGNKYSYRCNDGSPSPGSDYLRCTAQWEPVPCVRKCVFHVENGDS
AYWKIYVQGSGLKQCYNGYSLQNGODIMTCTENGWSPPKCIRIKTCSASDIHIDN
GFLSSSVIYALNRTSYRCQGYVNTGEISGSITCLQNGWSPPOPSIKSDMPVE
NSITKNTBWEKLDKDYECLVGFENEYKHTKGSITCTIYGNWSDTPSCYDSTRTCGP
PPIDNGDITSLPEYELPSVDYQCCYKYLKGNKTTICRNGKWSPPPTCLHACVI
PENMEAHNIIILKWHHTKIIYAHSGEDIEFECKRGYQKARGSLPRTKINGTINPT
CV"
misc_feature 876..1027
/note="intron-like sequence"
exon 1028..1204
misc_feature 1205..1364
/note="intron-like sequence"
exon 1365..1744
/note="complement factor H-related protein"
BASE COUNT 779 a 430 c 409 g 738 t
ORIGIN Chromosome 1.

Query Match 14.1%; Score 594.6; DB 10; Length 2356;
Best Local Similarity 65.2%; Pred. No. 3.4e-122;
Matches 1002; Conservative 0; Mismatches 369; Indels 166; Gaps 2;

QY 793 ctggatgaatcctcagcctcctgtgaagaatgacatggttgactccatatatccaa 852
DB 117 CTGATGCGCTTCTACTGCTAAAGGGGAGAAAGACATGCTCACCCTCTTATCTAA 176

QY 853 atggtatctacacacctcacaggtataaacacagaaattgatgatgaatcagatgaat 912
DB 177 ATGGTATCTACACCTCACAGGATATACACAGAGTGATGATGAATCAGATATGAAT 236

QY 913 gtaaaaatggctctctatctcgaaccgatccctgtttccaaagtacaaattactggct 972
DB 237 GTAATATGGCTTCTATCTGCTGTAACCTGGATCAACTGTTTCAAGGTGTACACCCTGGCT 296

QY 973 ggatccctcctcaagatgtagctgaaacctgtgattttccacaattcaaaatggac 1032
DB 297 GATCCCTGTTCCAGAGTGACTTGAACCTTGTAATTTCCACAAATTCAAATATGGAC 356

QY 1033 gtctgtattatgaagaacggcgagaccctacttcccagtaacctataggaaggagtagca 1092
DB 357 GTCTGTATTATGACAGAGCCTGACACCACTTCCCAGTATCTATAGGAATAGTACA 416

QY 1093 gctataactgtgacaacgggtttacaacgcttcacagtcactcactcagtcactcctcgtt 1152
DB 417 GCTATAGGTGTGACAAACGGGTTTCCACCCTTCTGGGTATTCTCTGGGACTACCTTCGT 476

QY 1153 gcacagtaaatgggtggagcctgaagttccatcctcaggaatgattttccattatg 1212
DB 477 GCACAGCACAGGGTGGGAGCTGAAGTCCCATGCTCAGGAAATGTTTTCATTATG 536

QY 1213 tggaaatggagaatcttcatactggaagaagaatataatagaggtcagtcctgcaaaag 1272
DB 1213 tggaaatggagaatcttcatactggaagaagaatataatagaggtcagtcctgcaaaag 1272

DB 537 TGGAGAAATGGAGACACTCTGCATACTGGGAAAAAATATATATGTGCAGGCTCAGTCTTTAAAG 596
QY 1273 tccagtgctcagtgctctatagctctccaaatgggtcaaacatcatattattgtacagaga 1332
DB 597 TCCAGTGTTCACAAATGGCTATAGTCTTCAAAATGGTCAAGACATAATGACATGTACAGAGA 656
QY 1333 atggtctgtccctcctcccaaatcgctccgtatccatcgaagacttggttcagtatcagatatag 1392
DB 657 ATGGCTGGTCCCTCTCCCAAAATGCGTATCAAGACATGTTCCAGCATCAGATATAC 716
QY 1393 aaatgaaaaatgggttttttctgaatcgtattatatacatatcgtctctaaalagaaaaaacac 1452
DB 717 ACATTGACAAATGGATTTCTTCTGAATCTTCTATATATATATGCTCTAAATAGAGAAACAT 776
QY 1453 ggtatagatgtaaacaggatattgtaacaaatcccgagagaaatatacaggaaataactt 1512
DB 777 CCTATAGATGTAAAGCAGGCGATATGTACAAATACTCGAGAAATATCAGGATCAATACTT 836
QY 1513 gttctaaagatggatgggtgcacctcgacctcatgcat- 1551
DB 837 GCCTTCAAAATGGATGGTGCACCTCAACCTCATCGCATTTATGAAGAAAGTCTTTGAAGAGCT 896
QY 1552 ----- 1551
DB 897 GCAACTCCCTGAAAGTGTGATGACCTTATAATATGTTTACACTGTGAGTTAAGGTCC 956
QY 1552 ----- 1551
DB 957 ATGTCAGAACATCTCCCTCTTAGAAGATCTTTTAAATATTTTATATTTGGAAGAAAGA 1016
QY 1552 -----agcttgatagatgcctgatttgagaattctatgactaagaataataa 1600
DB 1017 AATGCTTCCAGAGTCTGTGATATGCTGATTTGGAATTTCTATTAAGAAATCTACTAG 1076
QY 1601 cacatggtttaaacactcaatgacaaattagactatgaatgacattgacattggaatgaaataa 1660
DB 1077 GACATGGTTTAAAGCTCAATGACAAATTAGACTATGATGCTGCTGGTGGATTTGAAATGA 1136
QY 1661 atataacataccaaaggctctataacatgacttactatgatgagtgctctagtagcacccctc 1720
DB 1137 ATATAACATACCAAGGCTCTTAACATGTACTTATTATGATGATGCTGTGATACACCTTC 1196
QY 1721 ctggttatgaagaagaatgcagcattccctcgtttacaccaagacttagtggtttttccag 1780
DB 1197 ATGTTATGTTCTCCATCCCTCTACCTCTGTGATCATTTTGTTCGCCCTTCTAAGTGG 1256
QY 1781 aagaataaaatacaaaatggagattcgttgagtttctctcgttcaggaacacagagt 1840
DB 1257 GATTGAAGCATATACACTTTTGACACTTTGGTCTTACTTCTTTTAAAGCATCATATGGTCT 1316
QY 1841 tggagcagatttagtgcaatgctacacactttggatgggtcccttaatttcccaacggtgga 1900
DB 1317 ATGAATTTGATTGGTATTTCTGAGCTTTTGGGATATCTACATATCAACTCAACAAGCA 1376
QY 1901 egggcaagtataaatcatgtgacacacccctcttgaaatcccgaaatggggaaataaagggaac 1960
DB 1377 -----CATGTGGGCCCTCTCCACCTATTGACAAATGGAGACATCACCTCTT 1422
QY 1961 aaaaaagttgaatcacagccatgggagcgtggtggaatggaatgattgacacactagattct 2020
DB 1423 GTCATTACAGANTATGAAACCATTTATCATCAGTTGACTATCAGTGCCTGACCAAGATATTATCT 1482
QY 2021 actgaagggaacccaataaaatccagtggttgacgggaagtggacaaggtgtcccgatag 2080
DB 1483 CCTTAAGGGAATAAGACAAATAACATGTAGAATGGAAAGTGGTCTGAGCCACCAACGTG 1542
QY 2081 cgttgatgatgagagaacatgtggagaccttccctgaacttgagcattggtctgtcaagtt 2140
DB 1543 CTTACATGCATGTGTAATACCAAGAAACATTTATGGAAGCACACAATAATAATTTCTCAATG 1602
QY 2141 atctgtccctccctaccatcatcattgagattcagtgagag 2177
DB 1603 GAGACACACTGAAAAGATTTTATGCCCCATTCAGGGGAG 1639

```

RESULT 15
BC012610
LOCUS
DEFINITION Homo sapiens, clone IMAGE:39962233, mRNA.
ACCESSION BC012610
VERSION BC012610.1 GI:15706463
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 1486)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (15-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbiology.org
contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 19 Row: a Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4504374.

FEATURES
source
1..1486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:39962233"
/tissue_type="Bladder, carcinoma"
/clone_lib="NIH_MGC_53"
/lab_host="DH10B"
/notes="Vector: pDNR-LIB"

BASE COUNT 471 a 270 c 331 g 414 t
ORIGIN

Query Match 11.7%; Score 493.2; DB 9; Length 1486;
Best Local Similarity 64.1%; Pred. No. 1.4e-99;
Matches 874; Conservative 0; Mismatches 298; Indels 192; Gaps 1;

QY 5 gtcaactgctcccgatagatccaagaagctgagactgtcagcaagaattatttgcttat 64
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 65 attatggaactgtttgtatgacagaagattgaaggctctctctccagagaaaattcaga 124
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 99 GTTATGGGCTATTGTGTACGAGAAGATTGCAATGAACCTTCTCCCAAGAAGAAATACAGA 158
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 125 aattctctcagggttcgggtctgaacaactatttcagaaggcactcagcaacctacaa 184
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 159 AATCTGCACAGGTTCTGGTCTGACCAACATATCCAGAAGGCCACCCAGGCTATCTATAA 218
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 185 atgccgcccttgataccgaacaacttggtgactattgtaaaaagtgatgaagaattggagaatg 244
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 219 ATGCCGCCCTGGATATAGATCTCTTGCAAAATGTAATAATGGTATGCAGGAAGGAGGAATC 278
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 245 ggtacctcttaaccctcaaggatattgcgaaaaggaatgtgggcattccccggagacac 304
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 279 GGTGTGCTTTAAATCAATPAAAGGAAATGTCAAGAAAGCCCTGTGGACATCTCTGGAGATC 338
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Search completed: August 30, 2002, 19:06:49
Job time: 11894 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 14:57:15 ; Search time 3911.83 Seconds
(without alignments)
14591.284 Million cell updates/sec

Title: US-09-316-163-1
Perfect score: 4229
Sequence: 1 tcagatcaactgctcccgca.....aaaaaaaaaaaaaaaaaaaaa 4229

Scoring table: IDENTITY_NUC
Gapop 10.0 , Capext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estto: *
8: em_estc: *
9: gb_estli: *
10: gb_est2: *
11: gb_htc: *
12: gb_gss: *
13: em_gss_hum: *
14: em_gss_inv: *
15: em_gss_pln: *
16: em_gss_vrt: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	671.8	15.9	1113	10	BM461070
2	583.4	13.8	881	9	AA833130
3	558.8	13.2	903	9	AL531827
4	558.2	13.2	818	9	AI006101
5	554.6	13.1	906	9	AL540473
6	544.6	12.9	631	9	AI170314
7	538	12.7	626	9	AI009773
8	534	12.6	726	10	BI327605
9	520.6	12.3	548	10	BM383739
10	517	12.2	665	9	BB617140
11	514.4	12.2	661	9	AW701301
12	512.2	12.1	652	10	BF168914
13	510.6	12.1	842	9	AI122695
14	505.2	11.9	759	9	AA882470
15	496.6	11.7	652	9	BB609000
16	494.4	11.7	891	10	BF237071
17	494	11.7	595	10	BF395144

ALIGNMENTS

RESULT 1

BM461070 1113 bp mRNA linear EST 05-FEB-2002

LOCUS AGENCOURT_6419764 NCI_CGAP_OV44 Mus musculus cDNA clone

DEFINITION IMAGE:5503808 5', mRNA sequence.

ACCESSION BM461070.1 GI:18510110

VERSION BM461070

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE NIH-MGC http://mgc.nci.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Aaron Hsueh

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution by: Agencourt Bioscience Corporation

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

High quality sequence stop: 662.

plate: LLAM2144 row: c column: 09

Location/Qualifiers

1. .1113

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:5503808"

/clone_lib="NCI_CGAP_OV44"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: ovary, PMSG-treated; Vector:

pcMV-SPORT6.ccd; Site:1: EcoRV; Site:2: NotI; Cloned

unidirectionally. Primer: Oligo dT. Average insert size

2-2 kb. Library constructed by Life Technologies. Note:

this is a NCI_CGAP Library."

BASE COUNT 353 a 224 c 235 g 300 t 1 others


```
Db 679 ATCCCCAGATGTTATAAATGGATCTCCTATATCTCAGAAGATTATTATATAAGGAGATGA 738
QY 725 aagattccaataaataaggaaggttttgggtacaaagaaagggaggtgctgtctcg 784
Db 739 ACCATTTCAATATTAATGTAACATGGGTATGATGAATACAGTGAAGAKGAKATCTATATK 798
QY 785 caggggttctggatgaatctctcagcttctctgtggaagaaatgacatgtttgactccata 844
Db 799 CACTGAATCTGGATGCGCTCGCTTGCCTTCATGTGAAGAAAATCATGTGATAATCCTTA 858
QY 845 tattccaaatggtatctacacacctcacaggattaa 880
Db 859 TATTCCAAATKGCTACTACTACCTTTAAGGATAAA 894

RESULT 4
AI006101 818 bp mRNA linear EST 12-JUN-1998
LOCUS ua86f01.r1 Soares_mammary_gland_NBMWG Mus musculus cDNA clone
DEFINITION IMAGE:1364377 5' similar to gb:Y00716 COMPLEMENT FACTOR H PRECURSOR
(HUMAN); gb:M12660 Mouse CFH locus, complement protein H gene,
complete cds, (MOUSE);, mRNA sequence.
ACCESSION AI006101
VERSION AI006101.1 GI:3215710
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 818)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 505.
FEATURES
Location/Qualifiers
1..818
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1364377"
/clone_lib="Soares_mammary_gland_NBMWG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7m3D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(df)
primer [5
TGTACCAATCTGAAGTGGGCGCGCGCAATGGTTTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
252 a 161 c 196 g 209 t

BASE COUNT
```

ORIGIN

```
Query Match 13.2%; Score 558.2; DB 9; Length 818;
Best Local Similarity 85.4%; Pred. No. 1.6e-88;
Matches 645; Conservative 0; Mismatches 108; Indels 2; Gaps 2;

QY 2673 ccaggtctccaaagagagagatttaattgagtcacagcagttatgaacgcggaact 2732
Db 15 CCCAGATCTTCAGAGAAAGGAGAGATTCCATTGAGTCCAGCAGTCATGAACATGGAAC 74
QY 2733 acattcagctattgctgtgagagatggattcaagatcatctgaagaataatgggtaacctgc 2792
Db 75 ACATTCAGCTATGCTGCTGATGATGCTTTCAGATACCTTCCAGAGAAATAGATAAACCTGC 134
QY 2793 aacatgggaaatggagctctgctctgctgtgtgtgtggaatacctctgtgagcccccaact 2852
Db 135 TACATGGGAAAATGGAGCACCTCCACCTCGCTGTGTGGACCTTCCTTGTGGACCTCCACCT 194
QY 2853 tcaattcctctgtgattgtttctcatgaactagaagaattaccataatgatggagaggggtt 2912
Db 195 TCAATTCCTCTTGGTACTGTTCTCTTGACCTAGAGAGTTACCAACATGGGGAAGAGTT 254
QY 2913 acatacaattgttctgaagcgtttggaattgatggaccagcatttattaaatgttagga 2972
Db 255 ACATACCATTGTTCTACAGGCTTTGGAATTTGATGGACCAGCATTTATTATATGCGAAGGA 314
QY 2973 gacagtgctgtaaacctccccaaatgcataaaaactgattgtgacaacttgcccacattt 3032
Db 315 GGAAAGTGTGTCACCCACCAAAATGCATAAAACGGATTTGTGACGCTTTTACCCACAGTT 374
QY 3033 gaaattgccaaacccgacagaaagaaataatcacaaggttcaggagaaacaagtga 3092
Db 375 AAAAATGCCATAATAAGAGAGAAAGAGACAAATAATCATATAGGACAGAGACAAGTGACA 434
QY 3093 ttcagatgtccactccgtatcgaaatggatgctgcagactgtgcacatgtgttaatacg 3152
Db 435 TTCAGATGTCAATCTCCTTATCAAAATGAATGGCTCAGACACTGTGACATGTGTTAATAGT 494
QY 3153 aagtgagattggacgcggtatgcaagaataattcctgtgtgaatccacacacatgtgcca 3212
Db 495 CGGTGGATTGGACAGCCAGTATGCAAAAGATAATTCCTGTGTGGATCCACACATGTGCCA 554
QY 3213 aatgctactataactaaagggcacaagactaaatccatctcgtgacaaagtacgttat 3272
Db 555 AATGCTACTATAGTACAAGGACCAAGATAAATAATCTACATGTCACAGAGTACCTTAT 614
QY 3273 gactgtaataaaccttttgaattatttggggaagtggagtgatgtgccaaaaacgggatt 3332
Db 615 GAATGTAATAAAGCTTTGGAACATAATTTGGGCCAGTGGAAAGTGATGTGAAAAATGGGATA 674
QY 3333 tggacagaaacccgaaatgcaagattccaacgggaaatgtggcctcctccacatt 3392
Db 675 TGGACCAGAAAACC-AAGTGTGAGACTCAGCGGGGAAATGTGGG-CTCCTCCACCTATT 732
QY 3393 gacaatggagacatcacctccttctgtcattaccagt 3427
Db 733 GACCATGGAGCATCACCTCCTGCTTACCGGT 767
```

RESULT 5

```
AL540473
LOCUS AL540473 LTI_FL002_PL1 Homo sapiens cDNA clone CS0DE001YB21 5 prime
DEFINITION ' mRNA sequence.
ACCESSION AL540473
VERSION AL540473.1 GI:12870654
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 906)
```

AUTHORS Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..906
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DE001YB21"
/clone_lib="LRI FL002_PL1"
/lab_host="DH10B"
/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 299 a 138 c 214 g 255 t
ORIGIN
Query Match 13.1%; Score 554.6; DB 9; Length 906;
Best Local Similarity 77.8%; Pred. No. 6.e-88;
Matches 682; Conservative 0; Mismatches 194; Indels 1; Gaps 1:
QY 5 gTcaactgtcccaagatagatccagacatgagactgtcagcaagaattatttgcttat 64
Db 30 GTAATGTCCTCTTAAAGATCAAAAATGAGACTTCTAGCAAGATATTGTGCTTAT 89
QY 65 attatggactgttgttagcagaagattgttaaagttctctcccaagagaattcaga 124
Db 90 GTATGGGCTATTGTGAGCAGAGATTGCAATGAACCTTCTCCAGAGAAATACAGA 149
QY 125 aattctcactggtgtgtgtgaacaactattcagaaggcactcaggcaacctacaa 184
Db 150 AATTCTGACAGGTTCCTGCTGACCAACATATCCAGAAAGCCACCCAGGCTATCTATAA 209
QY 185 atgcccctggtgataccgaacacttggtactattgtataaagtatcagaagaatggagaatg 244
Db 210 ATGCCGCCCTGGATATAGATCTCTTGGAAATGTAATAATGGTATGATGAGAGGACAATG 269
QY 245 ggtaccttctaaccatcaaggatgtgtcgaaaggccatgtggcattcccggagagcac 304
Db 270 GGTGCTCTTATCCATTAAAGGAATGTGAGAAAGGCCCTGTGGACATCTCGGAGATAC 329
QY 305 acccttgggtccctttagctggcagtggtggtatctgaatttgatgtgcaaaaggttgt 364
Db 330 TCCTTTTGTACTTTTACCCTTACAGGAGGAATCTGTTTGAATATGTTGTAAGAGCTGT 389
QY 365 ttatacatgtgatgaagggtaccactattagtgaaattgattccctgtaagtgtatgc 424
Db 390 GTATACATGTAATGAGGGGTATCAATTGCTAGGTGAGATTAATTACCGTGAATGTGACAC 449
QY 425 agatgggtggaccatgatattcccaatgtgaaagtgtgaaagtgtgcaagtgtgcaagtaacaga 484
Db 450 AGATGGATGGACCAATGATATTCCTATATGTGAAGTTGTGAAGTTTACCAAGTGACAGC 509
QY 485 actggagaatggaagaattgtgagtggtgagccgcaaacacagaccagcaataatttttgg 544
Db 510 ACCAGAGAATGAAAATTTGTGAGTGTGCAATGCAATGCAATGCAATGCAATGCAATGCA 569
QY 545 acaaggtgtacgtttgatacgaactccggttcaagattgaagagcaagaagaatgca 604
Db 570 ACAAGCAGTACGGTTGTGTGATGATCGAGGTACAAAGATTGAAGGAGATGAAGAATGCA 629
QY 605 ctgctcaaaaatggccctctggagcaatgaaagccacagtggtgtggaaattcttgcct 664

Db 630 TTGTTTCAGACGATGGTTTTTGGAGTAAAGAGAAACCAAGTGTGTGAAATTTTCATGCAA 689
QY 665 gcccacacgagttgaaatggagatggtatatctgaaacacagtttacaagagaatga 724
Db 690 ATCCCCAGATGTTATTAATGGATCTCTATATCTCAGAGATTTATTATANGAGAGATGA 749
QY 725 aagattcccaataataatgtaagaaggtttttgtgtacaaagaagaaggaggtgctgtctg 784
Db 750 AGCATTTCAATATAAATGTAACATGGTTTATGAATACAGTGAAGAGAGAGATCTCTATG 809
QY 785 cagggttctggaatgaatccctcagccttccctgtgaaagaataagacatgtttg-actccat 843
Db 810 CACTGAATCTGGATGGCGTTCCTTCATGTTGAAGAAAATCATCTGTCAGTAAATCCCTT 869
QY 844 atattcccaatggtatctacacacccacagattaa 880
Db 870 ATATTCCAATGGTGACTACTACCTTTAAGGATTAA 906
RESULT 6
AI170314/c
LOCUS AI170314 631 bp mRNA linear EST 20-JAN-1999
DEFINITION EST216240 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
RLUG82 3' end, mRNA sequence.
ACCESSION AI170314
VERSION AI170314.1 GI:3710354
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
REFERENCE 1 (bases 1 to 631)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (REEST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Other_ESTs: TC52298
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
FEATURES
Location/Qualifiers
1..631
/organism="Rattus sp."
/db_xref="ATCC (inhost):2027791"
/db_xref="taxon:10118"
/clone="RLUG82"
/clone_lib="Normalized rat lung, Bento Soares"
/note="Organ: lung; Vector: pT73Pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT 212 a 97 c 104 g 218 t
ORIGIN
Query Match 12.9%; Score 544.6; DB 9; Length 631;
Best Local Similarity 96.0%; Pred. No. 4e-86;
Matches 606; Conservative 0; Mismatches 9; Indels 16; Gaps 4:
QY 3588 ctccagatggaggaatgcaaatgtttattcccaatcaggagagaattggaattcatg 3647
Db 631 CTCAGATGAGGGAATGCAAGATTTATCCCAATCAGGGGAGAAATTTGAATTCATG 572
QY 3648 tgaacacctggatagaaaattcagagatcacctccgttctcgtaaacagtgcatgag 3707
Db 571 TGTAAACCTGGATATAGAAAATTCAGAGGATCACCTCCGTTTCGTACAAGTGCATTGAG 512
QY 3708 ggtcacatcaattatcccaactgtgtataaaaatcgatatcaaatattagtaaaccttat 3767

```
|||||
Db 511 GGTACATCAATTATCCACTGTGTATAAAATCGTATACAATTTATTAGTAAACCTTAT 452
QY 3768 ggaag- - - - - aaaaatgcacgtatattactaataacacagtttgaatttacattt 3816
Db 451 GGATGAACCTTTGTTTGAATAATGACATGTATATTACTAATACAGTTTGAATTTTACATTT 392
QY 3817 -aaatatgttttagctcttctcttaataagatatataaaactttttttatatlggtggtla 3875
Db 391 GAAATATTGTTTAGCTCATTTCTTCTAATAAGTATATAAACTTTTATTATATGTTGGTTA 332
QY 3876 atcagaactttacagactgttgcacaaagcaagaacattacattcacaacclccaatc 3935
Db 331 ATCAGTAACTTTTACAGACTGTTGCCACAAAGCAGAACATTGCAATTCAAAACCTCCTAATC 272
QY 3936 c-aaatatgatgtcccaagacaaactatgtctaaagcaagaataaataattgaattctt 3994
Db 271 CAAATATGATATGTTCCAAAGCAACAATATGTTCTAAGCAAGAAATAAATGTTAGTCTT 212
QY 3995 caatgtctgtttttattcaggaccctttcagattttcttggatacctt- - - ttgttaggtt 4051
Db 211 CAATGTCGTGTTTATTTCAGGACTTTTTCAGATTTTCTTGGATACCTTTTGTGTTAGGTT 152
QY 4052 ctgattcacagtgaatgaagacacactgaactgaacttcaaatgaattagttacttgcatt 4111
Db 151 CGATTTCAGTGTGAGGAGACACACTGACTCTGACTTCAAAATGATTACTTTGCCAA 92
QY 4112 acattaaacacaaactatcatataatcacaaatgtatcacagtaattactgtgtcctac 4171
Db 91 TACATAACAACTTATCATATATATACAAATGATATACAGCTTAATTACTGTGTCCTAC 32
QY 4172 cttgtatcaataaagaatacttaagaagtt 4202
Db 31 CTTTGTATCAATAAAGAAATCTAAGAAAGTT 1

RESULT 7
LOCUS AI009773/c 626 bp mRNA linear EST 25-JAN-1999
DEFINITION EST204224 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
ACCESSION R1UB029.3' end, mRNA sequence.
VERSION AI009773
KEYWORDS AI009773.1 GI:3223605
SOURCE EST.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R. and Adams, M.D.
Rat Genome Project: Generation of a Rat EST (RAT) Catalog & Rat
Gene Index
Unpublished (1998)
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3525
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1..626
/organism="Rattus sp."
/db_xref="ATCC (inhost):2020517"
/db_xref="taxon:10118"
/clone="R1UB029"
/clone.lib="Normalized rat lung, Bento Soares"
/notes="Organ: lung; Vector: pT7T3pac; Site_1: EcoRI;
Site_2: NotI."
BASE COUNT 213 a 95 c 103 g 215 t
ORIGIN
```

```
Query Match 12.7%; Score 538; DB 9; Length 626;
Best Local Similarity 95.8%; Pred. No. 5.9e-85;
Matches 600; Conservative 0; Mismatches 10; Indels 16; Gaps 4;

QY 3571 aaaaacataataagttctcagatggagggaataagcaagatttattcccaatcagggg 3630
Db 626 AAAAAACATAATAGTCTCAGATGGAGGGAATAATGCAAGATTATTCCCAATCAGGGG 567
QY 3631 agaattatgaattcattgtgttaaacctggatatagaaaaattcagagatcacctccglttc 3690
Db 566 AGAATATTGAATTCATGTGTAAACCTGGATATAGAAAATTCAGAGGATCACCTCCGTTTC 507
QY 3691 gtacaaagtgaattgaggggtcacatcaattatcccaactgtgtataaaaatcactatacaa 3750
Db 506 GTACAAAGTGCAATGAGGGTCCATCAATTTATCCCACTGTGTATAAAAATCCGTATACAA 447
QY 3751 ttattagtaaaccttatggatg- - - - - aaaaatgcacatgtatatlaactaatc 3799
Db 446 TTATTAGTAAACCTTATGGATGAACCTTTGTTTAGAAATGCACATGTATATTACTTAATAC 387
QY 3800 agtttgaattacattt-aaatatgttttagctcttctctcttaataagatatataaactt 3858
Db 386 AGTTTGAATTTACATTTGAAATATTGTTTAGCTCATTTTCTTAATAAGTATATAAATTT 327
QY 3859 tttttatatgtgtgttaataatcagtaactttcagactgttgcacaaagcaagaacattac 3918
Db 326 TTTTATATGTTGGTTTAAATCAGTAATTTACAGACTGTTGCCACAAACCAAGCAATGCC 267
QY 3919 attcaaaactcctaatcc-aaatatgtatgtcccaagacaaactatgtctaaagcaagaa 3977
Db 266 ATTCAAAACTCCTAATCCAAAATATGATATGCCAAGGACAAACTATGTCTAAGCAAGAA 207
QY 3978 aataaatttagttcttccaatgtctgtttttatttattcaggacccttcagattttcttggata 4037
Db 206 AATAAATGTTAGTCTCTCAATGTCGTTTTATTATCAGGACTTTTTCAGATTTTCTTGGATA 147
QY 4038 cctt- - - ttgttagttcttctcagttcagttgagtggaagacacactgactctgaactcaaa 4094
Db 146 CCTTTGTTGTTAGGTTTGTGTTTACAGCTGAGTGGGAAGACACACTGCTGACTCTCAAA 87
QY 4095 ttagtattactgcaataacatacaacaaactatcatataatcacaaatgtatacacgc 4154
Db 86 TTAGTATTACTTGTCCCAATACATAACAACCAAACTATCATATAATACAAAATGTATACAGC 27
QY 4155 taattactgtctcctaccttctgtatc 4180
Db 26 TAATTACTGTGTCTTACCCTTTGTATC 1

RESULT 8
LOCUS BI327605 726 bp mRNA linear EST 30-JUL-2001
DEFINITION 602979789F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5132471 5',
mRNA sequence.
ACCESSION BI327605
VERSION BI327605.1 GI:15012262
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
```



```

|||||
Db 242 CAATGAGACATCCTCTCTCATTACCATGATATGACCATATATCATCATGTCGAATA 301
QY 3455 tcaatgccagaactattatctacttaagggaaataagatagtaacatgtagaatggaaa 3514
Db 302 TCAATGCCAGAAGTATTCTCTCTTAAGGGAAGAAGACAATAACATGTAGAAATGGAAA 361
QY 3515 gtggctcagccaccacactgtctacatgcatgtgtgataccagaagatattatggaaaa 3574
Db 362 GTGGCTGAGCCACCACCATGCTTACATGCATGTGTAAATACCGAGAAACATTATGGAATC 421
QY 3575 acataatagttctcagatgaggggaaaaatgcaagattttatcccaatcaggggagaa 3634
Db 422 ACACAATATAAATCTCAAAATGAGACACACTGAAAGATTTATTCCTCCATTCAGGGAGGA 481
QY 3635 tattgaattcatgtataaacctggatatagaaaaattcagagatcacacctccctttogtac 3694
Db 482 TATTCAATTTGGATGTAATATGGAATATATAAGCAAGAGATTTACCCGCCATTTTCGTAC 541
QY 3695 aaagtgcattgaggtcacatcaattatccactgtgtataaaatcgctatacaattat 3754
Db 542 AAAGTGCAATATGCGACCATCAATATATCCCACTTGTGTATAAAATCAATATACATTTAT 601
QY 3755 tagtaacctatgtagagaaatgcacatgtattactataatacagtttggattacat 3814
Db 602 TAGTTGATTNTATGCTTAGAAAGGCACATGATGCTGACTAATATATCTNTCAATTTGCAT 661
QY 3815 tta 3817
Db 662 TGA 664

```

```

RESULT 11
AW701301/c
LOCUS
DEFINITION
IMAGE:2537283 3' similar to gb:M12660 Mouse CFh locus, complement
protein H gene, complete cds, (MOUSE);, mRNA sequence.
ACCESSION
AW701301
VERSION
AW701301.1 GI:7585430
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 661)
AUTHORS
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Willson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
CONTACT: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL : contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:019879
Seq primer: custom primer used
High quality sequence stop: 501.
Location/Qualifiers
1..661
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2537283"
/clone_lib="Sugano mouse kidney mkia"
/sex="female"
/dev_stage="adult"

```

FEATURES

```

source
1..661
/lab_host="DH108"
/note="organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
(CACTGTGG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCTTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCTCTCTCTAAAGCTGCC and 3' end
primer CGACCTGCAGCTCGACACA."
BASE COUNT 191 a 136 c 127 g 207 t
ORIGIN
Query Match 12.2%; Score 514.4; DB 9; Length 661;
Best Local Similarity 86.2%; Pred. No. 8.3e-81;
Matches 569; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
QY 1099 actgtgacacaggggttacacgcttcacagtcacactgagggactacacctcgttgcacag 1158
Db 661 AGTGTGACAAACGGGTTTTCACCACATTTCTGGGTATTTCATGAGACTACCTTCGTGTGCACAG 602
QY 1159 taaatgggtggagcctgaagttccatgcctcaggaatgtatttccattatgtggaat 1218
Db 601 CACAAGGGTGGGAGCCTGAAAGTCCCATGCTGCAGGAAATGTGTTTCCATTATGTGGAGA 542
QY 1219 atggagaatcttcatactactggcgaagaagatatataagagggtcagctgcaaaagtcagct 1278
Db 541 ATGGAGACTCTGCATCTACTCTGGAAGTGCATATGTCGAGGGTTCAGTCTTTAAAGATCCAGT 482
QY 1279 gtcaacagtggtctatagctctccaaatggcgaagatacatattattgcagagaatggct 1338
Db 481 GTTACAATGGCTGTAGTCTTCAAAATGGTCAAGACACAATGACATGATACAGAGAATGGCT 422
QY 1339 ggtccctcctcccaaatgcgtccgtatcaagactgttccagtatcagatatagaaattg 1398
Db 421 GGTCCCTCTCTCCAAATGCATCCGTATCAAGACATGTTTCAGCATCAGATATACACATTG 362
QY 1399 aaaaatgggttttttctgaatcgtgattacatatatgctctaaataaagaaacacggata 1458
Db 361 ACAATGGATTTCTTCTGAAATCTTCTTATATATGCTCTAAATAGAGAAACATCTCTATA 302
QY 1459 gatgtaaacagggatattgtacaataaccggagaaatatcaggaaataattactgtcttc 1518
Db 301 GATGTAAACAGAGGATATGTGACAAATACTGGAGAAATATACAGATCAATAACTTGCCTTC 242
QY 1519 aagatggatggttcacctgcacctcctcattgaagctctgtgatatgcctgtatttgaga 1578
Db 241 AAAATGGATGGTTCAGCTCAACCTCTCATGCTTAAAGTCTTGTGATATGCTGTATTGAGA 182
QY 1579 attctactaataagaataaacacatggttttaactcaatgacaaattagactatgaat 1638
Db 181 ATCTTAACTAAGAATACTAGGACATGGTTTAAAGTCAATGACAAATAGACTATGAAT 122
QY 1639 gtcaattggatgataaaatgaataaataaccacaaaggtctataacatgactattg 1698
Db 121 GTCTCGTGGATTTCAAAATGAATATAACATACCAGGCTCTATAACATGCTACTATT 62
QY 1699 atggatggtctagtcacacctcctctgtatgaagaagaatgcagcattccctgttacacc 1758
Db 61 ATGGATGTTTGTATACACCCCTCATGTTATGAAAGAGAAATGCAGTGTGCCACTCTAGCCC 2

```

RESULT 12

```

BF168914
LOCUS
DEFINITION
601775377F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:401715 5',
mRNA sequence.
ACCESSION
BF168914
652 bp mRNA linear EST 30-OCT-2000

```

VERSION	BF168914.1	GI:11049266
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE	NIH-MGC http://mgc.nci.nih.gov/	
JOURNAL	1. (bases 1 to 652)	
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-re@mail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM9265 row: 1 column: 20 High quality sequence stop: 648.	
FEATURES	Location/Qualifiers	
source	1..652	
	/organism="Mus musculus"	
	/strain="CZECH II (feral)"	
	/db_xref="taxon:10090"	
	/clone="IMAGE:4017115"	
	/clone_lib="NCI-CCAP Lu29"	
	/tissue_type="spontaneous tumor, metastatic to mammary."	
	Stem cell origin."	
	/lab_host="DH10B"	
	/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"	
BASE COUNT	216 a 139 c 134 g 161 t	2 others
ORIGIN		
	Query Match 12.1%; Score 512.2; DB 10; Length 652;	
	Best Local Similarity 87.6%; Pred. No. 2e-80;	
	Matches 570; Conservative 0; Mismatches 80; Indels 1; Gaps 1;	
Qy	2414 aaatgggattcgaacaaactgtacaagcaaaagtcttgcctctctcccccgaagat	2473
Db	1 AAAATGGGATCCTGAAACCAACTGTACAAGCAAAACATCTCTGCCTCTCTCACGCGCAGAT	60
Qy	2474 tccaaatgccaaagtattgaaccaccgcgtgaaaactacttgatggagaaaaagtatcgt	2533
Db	61 TCCAAATACCCAAAGTCATTGAAACCAACCCCTGAAATACTTGGATGGAGAAAATTAATCTGT	120
Qy	2534 tctttgccaaagtatgtaactaaactcagggccgaagaagaaagtgtgtataacatggaag	2593
Db	121 TCTTTTCCCAAGACAATTACCTAACTCAGGACTCAAGAAATGTGTGCAAAAGATGGAAG	180
Qy	2594 gtggcagtcgttaccacgctgcacggagaaaaattccattgttcccgccctaaaaattga	2653
Db	181 GTGGCAGTCATTACCTCGCTGATTGAAAAAATTCCTATGTTCCAGGCCCTTACAATAGA	240
Qy	2654 acatgatctatataagtcgcccaggtctctcagaagagaggagatttaattgagtcag	2713
Db	241 ACATGGATCTATTAAATTACCCAGATCTTCAGAGAAAAGAGAGATTTCATTGAGTCCAG	300
Qy	2714 cagttatgaacaggaactacattcagctattgtctgtagatgattcaagatatcga	2773
Db	301 CAGTCATGAACATGGAACATACATTACGCTATGCTCTGTGATGATGGTTTCAGGATACCTGA	360
Qy	2774 agaaaatagggttaacctgacaactgggaaaaatggagctctctgcctcgttgtgtggaat	2833
Db	361 AGAAAAATAGGATAAAGCTGCTACATGGGAAATGGAGCACATCCACCTCGCTGTGTGGACT	420
Qy	2834 acctgtgaccccaaccttaattccctctcgttggtatgttttctcatgaaatgaagatta	2893

```

Db 226 ATGCGCCCTGGATATAGATCTCTTGGAAATATAATAATGATGATCGAGGAGGAGATG 205
QY 245 ggtaccttctaaaccatcaaggatgtcggaagccatgtgggcatcccgagagacac 304
Db 286 GGTGCTCTTAATGAAGAAATGATCGAGAAAGCCCTGTGGACATCCTGGAGATAC 345
QY 305 accttgggtctttaaggctggcagttggatctgaaatttgaatttgggtcgaagggtgt 364
Db 346 TCCTTTTGGTACCTTTACCCCTTACAGGAGGAAATGTGTTGAAATATGTTGTAAGCTGT 405
QY 365 ttatacatgtgatgaaggggtaccacactattagggtgaattgattaccgtgaattgtgatgc 424
Db 406 GTATACATGTAAATGAGGGGTATCAATTTGCTAGGTGAGATTAATACCGCTGAATGTGACAC 465
QY 425 agatgggtggaaccaatgatattcccaatatgtgaagtgtgaaagtgttgccttgcagtgacaga 484
Db 466 AGATGGATGGACCAATGATATCTCTATATGTGAAGTTGTGAAGTGTGTACCAAGTGCAGC 525
QY 485 actggagaatggaagaattgtgagtggtgcagccgaaccagacagaccagaataattatttgg 544
Db 526 ACCAGAGAATGGAAAATTTGTCAGTAGTGCATGGAACCAACCAAGTGTGTGCAATTTTCATGCAA 585
QY 545 acagggtgttacgcttgaatggaactccggcttcacagattgaaggacagagaagaaatgca 604
Db 586 ACAAGCAGTACGGTTTGTATGTAACTCAGGCTACAGATTGAAGGAGATCAAGAAATGCA 645
QY 605 ctgctcataaaatggcctctggagcaatgaaagccacagctgtgtggaatttcttgcct 664
Db 646 TTCTTCAGACGATGGTTTTTGGAGTAAAGAGAAACCAAGTGTGTGCAATTTTCATGCAA 705
QY 665 gccaccacaggttgaaatggagatggttatatctgaaaccagtttcaagagagaatga 724
Db 706 ATCCCGAGATGTTAATGGATCTCTATATCTCAGAGATTTATTAAGGAGAATGA 765
QY 725 aagattcccaataaataagcaaggttttgtgtacaaagaagggtgagctgtgtctgt 784
Db 766 ACGATTTCAATATAAATGTAACATGGGTATGAATAACNGTGAAGAGGANATCCTGNATG 825
QY 785 caccggttctgg 796
Db 826 CCTGGATCTGG 837

RESULT 14
AA882470
LOCUS
DEFINITION
  vx45b10.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone
  IMAGE:1278139 5' similar to gb:Y00716 COMPLEMENT FACTOR H PRECURSOR
  (HUMAN); gb:M12660 Mouse CFH locus, complement protein H gene,
  complete cds, (MOUSE);, mRNA sequence.
ACCESSION
  AA882470
VERSION
  AA882470.1 GI:2991581
KEYWORDS
  EST.
SOURCE
  house mouse.
ORGANISM
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 759)
REFERENCE
  Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
  Geisel,S., Kucaba,T., Lacy,M., Le.M., Martin,J., Morris,M.,
  Scheinberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
  Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilton,R. and
  Waterston,R.
  The WashU-HMI Mouse EST Project
  Unpublished (1996)
JOURNAL
  Contact: Marra M/Mouse EST Project
  WashU-HMI Mouse EST Project
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: mouseest@watson.wustl.edu

```

```

FEATURES
  source
    1..759
      /organism="Mus musculus"
      /strain="C57BL/6 x CBA"
      /db_xref="taxon:10090"
      /clone="IMAGE:1278139"
      /clone_lib="Stratagene mouse lung 937302"
      /sex="female"
      /tissue_type="lung"
      /dev_stage="6-8 month old"
      /lab_host="SOLR (kanamycin resistant)"
      /note="Organ: lung; Vector: pBluescript SK-; Site 1: ECORI
      ; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo
      dt. 6-8 month old female lung and 1.5 year old male lung
      were source of mRNA. Average insert size: 1.5 kb; Uni-ZAP
      XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'
      adaptor sequence: 5' CTCGAGTATTTTTTTTTTTT 3'"
BASE COUNT      244 a 168 c 153 g
ORIGIN
  QY 2118 ctgagcagtgctctgcaagttatctgtccctccctaccatcatgagagattcagtgag 2177
  Db 1 CTTGAAATGGCTGTGTCAAGTTTCTGTCTCCCTCCCTACCACCATGGAGATTCAGTGGAG 60
  QY 2178 ttcaactgtacagaacaccttcacaatgattggacatgagcagtagtttctcattagtgga 2237
  Db 61 TTCACTTGTGCAGAAACCTTCACAAATGATTTGGACTTGGGTGAGTCTTCTGCCCTTAGTGGGA 120
  QY 2238 aggtgacagagcttccctcctcaatgattgttgcaacagatcaactgagagaagtgttaaagccccc 2297
  Db 121 AAGTGGACCCAGCTTCTCTAAATGTGTGCAACAGACCAACTGGAGAAAGTGTAGAGTGTCTG 180
  QY 2298 aagtcacactggcagatgcaattcctccaaataagaatgaatttaatacatttagt 2357
  Db 181 AAGTCAACTGACATAGAAGCAATTAACAAAAGAAATGAATTTTCAGCAATACTCCACC 240
  QY 2358 gtgagttacagatgtagacaaaagcagagagatgataacattcaactctgcatcaatggaga 2417
  Db 241 ATGTATTACAAATGTAGAGACAAAGCAGGAGTATGAACACTCAATCTGTATCAATGGGAAA 300
  QY 2418 tgggatacttgaaacaaactgtacaaagcaaaagattctgcctcctcccccgcagattcca 2477
  Db 301 TGGAAATCTGAACCAAACTGTACAAGGAAAACATCTGCTCCCTCTCCACCACAGATTCCA 360
  QY 2478 aatgcccagaattgataaacaccggtgaaatacttggatggagagaaaaagtattgttctt 2537
  Db 361 AATACCCTAGTGTATGAAACCACTGTGAAATACTTGGATGGAGAAAATTTATCTGTCTTT 420
  QY 2538 tgcagaatgtgttaccataactcagggcccccagagaagaatggtgtgtaaacatggaagtgg 2597
  Db 421 TGCCAAGACAAATTACCTAATCTAGGACCCAGACAAATGATGTGCAAAAGATGGAAGGTGG 480
  QY 2598 cagctgttaccagctgcacgggaaaaattccatgttcccagccccctaaaaattgaacat 2657
  Db 481 CAGTCATTACCTCACTGCATTTGAAAAAATTCATGTTCAGGCCCTTAAAAATAGAACAT 540
  QY 2658 ggaatctattaaagtcgcccaaggctcctcagagaagagagagagatttaattagtcacagct 2717
  Db 541 GGATCTATTAAATTAACCCAGAACTTCTTGGGA-AAAGAGAGAAAATCCCATTTGANGTCAGCAGT 599
  QY 2718 tatgaacacggaactacattcagctatttctgtgagatggatggaatcagatatctgaaga 2777
  Db 600 CACGAACTGGAACTACATTCAGTATGCTGTGATGATGGGTTC-AGATATCTGAAGAA 658

```

This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:669939
 Seq primer: -28ml3 rev1 ET from Amersham
 High quality sequence stop: 439.

Location/Qualifiers
 1..759

Query Match 11.9%; Score 505.2; DB 9; Length 759;
 Best Local Similarity 85.3%; Pred. No. 3.4e-79;
 Matches 598; Conservative 0; Mismatches 99; Indels 4; Gaps 3;

Qy 2778 aatagggtacacgtggaacatgggaatggagctctctgc 2818
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 659 A--TGGGATACCTGCTACATGGGAGATGAAGCACTCCCC 697

RESULT 15

BB609000 652 bp mRNA linear EST 26-OCT-2001
LOCUS BB609000 RIKEN full-length enriched, 2 days pregnant adult female
DEFINITION ovary Mus musculus cDNA clone E330009L21 5', mRNA sequence.

BB609000
VERSION BB609000.2 GI:16451137

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 652)

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,H., Kouda
Okazaki,Y., Okido,T., Saito,K., Sakai,K., Sano,H., Sasaki
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muranatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)

TITLE

JOURNAL

COMMENT On Dec 6, 2000 this sequence version replaced gi:11564176.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Konno,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN,
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues
Location/Qualifiers
1. .652
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="E330009L21"

FEATURES

source

adult female ovary"
/sex="female"
/tissue_type="ovary"
/dev_stage="2 days pregnant adult"
/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCCGCAACTCGAGTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATCTCGAGTTAATAATTAATCCCTCCCTCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLIC 1."

BASE COUNT 185 a 120 c 170 g 177 t
ORIGIN

Query Match 11.7%; Score 496.6; DB 9; Length 652;
Best Local Similarity 92.2%; Pred. No. 1 ie-77;
Matches 523; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
Qy 1 tcgagtcacactgtccccagatagatcccaagacatgagactgtcagcagaatatttggc 60
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67 TGGAGTCAGTTGGTGGCCAGAGAATCCAAATTAATGAGACTGTCCAGAGAATATTATGGC 126
| | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 61 ttattattgagactgtttgtacacagaagattgaaaggtccctcccaagagaaaatt 120
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 127 TTATTATTGAGACTGTTGTGCAGCAGGAAGATTGTAAGGTCTCTCTCCCAAGAGAAAAAT 186
| | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 121 cagaaattctctcaggttctgtgctgaacaaactatatcagaaggcactcagcacaact 180
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187 CAGAAATCTCTCAGGCTCGTGGTCAGAACAACTATATCCAGAGGCCACCCAGGCTACT 246
| | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 181 acaaatgccgccttgatccgaacacttgatctgtctataatgtaaaagtatgcaagaatgag 240
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 247 ACAAAATCCGCCCTGGATACCGAACACTTGGCACTATTGTAAAAAGTATGCAAGATGAA 306
| | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 241 aatgggtacattctaaccccaagagatgtcgaaaaggcca,gtgggcattccggag 300
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 307 AATGGGTGGCGTCTAAACCCATCCAGGATATGTCGGAAAAACCCCTGTGGGCATCCCGGAG 366
| | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 301 acacaccccttgggtcctttaggctgagctgtgagatctgaatttgaatttgggtcacaagg 360
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 367 ACACACCCCTTGGGTCTCTTAGGCTGGCAGTTGGATCTCAATTTGAGTTGGTGCATAATG 426
| | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 361 ttgtttatacatgtgataaggggtaccacactattagtgaaattgataccgtaagtgtg 420
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 427 TTGTTTATACCTGTGATGATGGGTATCAACTATTAGGTGAAATTCATTACCGTGAATGTG 486
| | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 421 atcagaatggtagaccaatgatattccaatatgtgaattgtgaagtgtgcttgcagtga 480
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 487 GTCCAGATGGGTGGATCAATGATATTCACATGTGGAAGTTGTGAAGTGTCTACTCTGTGA 546
| | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 481 cagaactggagaatggagaattgtgagttggtgagtcagcgaaccagaccaggaatattatt 540
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 547 CAGAACTCGAGAATGGAAGAATTTGTAGTGGTGCAGCAGAGAACAACAGACAGGAATACCTAT 606
| | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 541 ttggacaggtgtgtacgctttgaaatgca 567
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 607 TTGCACAGGTGGTGGCGGTTTGAATGCA 633
| | | | | | | | | | | | | | | | | | | | | | | | | |

Search completed: August 30, 2002, 17:19:59
Job time: 8564 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 16:09:45 ; Search time 537.93 Seconds
(without alignments)
13497.725 Million cell updates/sec

Title: US-09-316-163-1
Perfect score: 4229
Sequence: 1 tcagtgcaactgtctccagc.....aaaaaaaaaaaaaaaaaaaa 4229

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. Is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	437.4	10.3	767	18	AAV02792 Human partial comp
2	432.4	10.2	525	21	AAA43501 Mouse secreted exp
3	427.2	10.1	649	18	AAV02790 Human partial comp
4	405	9.6	688	18	AAV02794 Clone PRBS3FH2910
5	376.8	8.9	581	18	AAV02791 Clone PRBS3FH410 C
6	334.4	7.9	550	18	AAV02795 Clone PRBS3FH2910
7	325.2	7.7	1293	21	AACT77947 Human cancer assoc
8	321.6	7.6	759	20	AAAX39793 Gastric cancer ass
9	307.2	7.3	532	18	AAV02793 Clone PRBS3FH2910

10	244.8	5.8	472	18	AAV02796 Human partial comp
11	196	4.6	2033	20	AAX34737 Human complement f
12	194.2	4.6	385	18	AAV02798 Clone p253FH2576 #
13	192.6	4.6	385	18	AAV02797 Clone p253FH2576 #
14	115.6	2.7	137	22	AAF31736 Novel rat gene fra
15	113.4	2.7	2177	9	AAN82402 B subunit of human
16	102.6	2.4	465	22	ABA42243 Human breast cell
17	102.6	2.4	465	22	ABA52669 Human foetal liver
18	102.6	2.4	465	22	ABA22457 Probe #923 for gen
19	102.6	2.4	465	22	AAK00926 Human brain expres
20	102.6	2.4	465	22	AAK26380 Human bone marrow
21	102.6	2.4	465	22	AAI11013 Probe #946 for gen
22	102.6	2.4	465	22	AAI32275 Probe #961 used to
23	102.6	2.4	465	22	AAI00933 Probe #924 used to
24	84.4	2.0	177	22	ABA64580 Human foetal liver
25	80	1.9	177	22	AAI44752 Probe #13438 used
26	80	1.9	156	22	ABA7391 Human breast cell
27	80	1.9	156	22	ABA65281 Human foetal liver
28	80	1.9	156	22	ABA32382 Probe #10848 for g
29	80	1.9	156	22	AAK13693 Human brain expres
30	80	1.9	156	22	AAK39433 Human bone marrow
31	80	1.9	156	22	AAI20246 Probe #10179 for g
32	80	1.9	156	22	AAI45448 Probe #14134 used
33	80	1.9	156	22	AAI05950 Probe #5941 used t
34	78	1.8	177	22	ABA65279 Human foetal liver
35	78	1.8	177	22	ABA32380 Probe #10846 for g
36	76.2	1.8	1929	22	AAAS06029 Angiotensin conver
37	74.2	1.8	335	16	AAT20816 Human gene signatu
38	71.4	1.7	177	22	ABA65280 Human foetal liver
39	71.4	1.7	177	22	ABA32381 Probe #10847 for g
40	67.2	1.6	422	22	ABA51919 Human foetal liver
41	67.2	1.6	422	22	AAI31529 Probe #215 used to
42	67.2	1.6	1194	22	AAAS06052 Angiotensin conver
43	66.2	1.6	494	21	AAC71094 Single nucleotide
44	66.2	1.6	494	21	AAC71100 Single nucleotide
45	66.2	1.6	494	21	AAC71115 Single nucleotide

ALIGNMENTS

RESULT	1
AAV02792	
ID	AAV02792 standard; DNA; 767 BP.
XX	AAV02792;
XX	AC
DT	27-APR-1998 (first entry)
XX	
DE	Human partial complement factor H cDNA fragment 2.
XX	
KW	Complement factor H; tumour associated antigen; renal cancer;
KW	urogenital cancer; medicament; modulator; SS.
OS	Homo sapiens.
XX	
FN	W09738136-A1.
XX	
PD	16-OCT-1997.
XX	
PF	09-APR-1997; 97WO-US05710.
XX	
PR	06-MAR-1997; 97US-0812481.
PR	09-APR-1996; 96US-0015083.
PR	09-APR-1996; 96US-0630048.
PR	06-MAR-1997; 97US-0038614.
PA	(BARO-) BARD DIAGNOSTIC SCI INC.
XX	
PI	Enfield DL, Hass GM, Kinders RJ;
XX	
DR	WPI; 1997-512742/47.
DR	P-PSDB; AAW39156.

XX	Treating or screening for cancer, e.g. renal or urogenital cancer -
PT	by modulating or detecting tumour associated human complement factor H
PR	H related antigen, or nucleic acid encoding it
XX	
PS	Example 6B; Fig 7A; 104pp; English.
XX	
CC	This partial cDNA sequence encodes a region of the human
CC	tumour-associated complement factor H (CFH). This sequence is used
CC	in the identification of DNA encoding complement factor H related
CC	proteins and antigens from clone PRBS3PH2910 (see AAU02793-02795). The
CC	detection of this protein and a CFH antigen can be used in screening or
CC	for the treatment of renal or urogenital cancer, e.g. bladder, cervical
CC	or prostate cancer. Agents that may modulate this antigen could be used
CC	in the manufacture of a medicament for the treatment of a tumour cell.
XX	
SQ	Sequence 767 BP; 257 A; 140 C; 179 G; 191 T; 0 other;
	Query Match 10.3%; Score 437.4; DB 18; Length 767;
	Best Local Similarity 73.1%; Pred. No. 5.5e-94;
	Matches 561; Conservative 0; Mismatches 206; Indels 0; Gaps 0;
QY	2862 ctctggattgtttcctgaactgaaagtaccacaataggagaggaggttacacataa 2921
DB	1 catggtgttcagctccacatgcagacagttcatcagtagtgagaagaagttcacgta 60
QY	2922 tgtctcaagcttctggaattgatgccacgacctatttaaattgtgtaggagacagtgtg 2981
DB	61 tgtttgaaggttttgaattgatggcgctgcgaattgcgaatgcttctaggagaaaatc 120
QY	2982 tctgaacctcccgaatgcataaaactgattgtgacaaacttgcgcacacttgaattgcc 3041
DB	121 tctcacctccatcatgcataaaacagattgtctcagtttaccttagctttgaaaaatg 180
QY	3042 aaaccgcagaaaagaaaaaacattacacaggttcaggagacaagtacattcagatgt 3101
DB	181 ataccatggagagagaaggaatgtgtcataagcgcggtgagcaagtgacttaccttgt 240
QY	3102 ccacctcgtatcgaatggatggctctgacatttcacatgtgttaacacgaagtggatt 3161
DB	241 gcaacattacaaaaatggatggagccagtaatgtaacctaatagcagatggaca 300
QY	3162 ggacagccggtatgcaaagataatctctgtgtgaatccaccacatgtgccaaatgctact 3221
DB	301 ggaaggccaacatgcagagacacctctgtgtgaatccgccacagtcacaaaatgcttat 360
QY	3222 ataactaagaagcacaagactaaaataccatctggtgacaaagtcagttaaactgtgat 3281
DB	361 atagtgcgagacagatgagttaatatccatctcgttgagagagtcagttatcaatgtagg 420
QY	3282 aaaccttttgaattattttggggaagtggaaagtgtgtgccaaaacgggatttgacagaa 3341
DB	421 agccctatgaaatggtttggggaTggaagaagtgtatgtttaatggaaaactggacgaa 480
QY	3342 ccaccgaaatgcaagattccaacgggaaaatgtgggcctctccacctattgacaaatgga 3401
DB	481 ccacctcaatgcaagaattctacaaggaaaatgtggggccccctccacctattgcaaaatg 540
QY	3402 gacatcacctcttcttattaccagtatatgcaccattatcatcagttgaatatcaatgc 3461
DB	541 gacattacttctccctgttcagtatatgtctccagcttccatcagttgagtaccaatgc 600
QY	3462 cagaactattactctaaggaaaataagaatagaatacattagaaaatggaaaatgtgtct 3521
DB	601 cagaactctcatcaactctgagggTaaacaagcagataacatgtagaataggacaatgttca 660
QY	3522 cagccaccaactcgtctacatgcatgtgtgataccagaagaatatatggaaaaacaataa 3581
DB	661 gaaccacaaaatgcttaccatccgtgtgttaattcccgagaaaattatggaataataaac 720
QY	3582 atagttctcagatggaggggaaaatgcaaaagattattccccaatcagg 3628

CC in the exemplification of the present invention.
 XX
 SQ Sequence 525 BP; 150 A; 101 C; 131 G; 142 T; 1 other;

Query Match 10.2%; Score 432.4; DB 21; Length 525;
 Best Local Similarity 92.5%; Pred. No. 7 4e-93;
 Matches 454; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 tcagatcaactgtccagatgacatccagacatgagactgctcagcaagaattatttggc 60
 DB 35 tggagtcagttgtccagagaagatccaaattattgagactgctcagcaagaattatttggc 94
 QY 61 ttatatttgagactgtttgtgacagaagattgtaaaagttcctcctcacaagaaaatt 120
 DB 95 ttatatttgagactgtttgtcagcagaagattgtaaaagttcctcctcacaagaaaatt 154
 QY 121 cagaattctcaggttctgtgtcgaacaactatattcagaagcactcaggaacact 180
 DB 155 cagaattctcaggttctgtgtcgaacaactatattcagaagcactcaggaacact 214
 QY 181 acaaatgcgccttgatccacacacttggcactattgtaaaagtatcagaagaatggag 240
 DB 215 acaaatgcgccttgatccacacacttggcactattgtaaaagtatcagaagaatggag 274
 QY 241 aatgggtaccttctaacccatcaagatatgtcgaaaaggccatgtgggcatcccgag 300
 DB 275 aatgggtggcgtctaacccatccagatatgtcgaaaaggccctgtgggcatcccgag 334
 QY 301 acacaccccttgggtctccttagctggcagttggatcgaatttgaatttggtgcaagg 360
 DB 335 acacaccccttgggtctccttagctggcagttggatcgaatttgaatttggtgcaagg 394
 QY 361 ttgtttatcatgtgatgaagggtaccacattattagggtgaattgattaccgtgaatgtg 420
 DB 395 ttgtttatcatgtgatgaagggtaccacattattagggtgaattgattaccgtgaatgtg 454
 QY 421 atgcagatgggtggaccatgatattccaatatgtgaaagttgtaagttgacagtgga 480
 DB 455 gtcagatgggtggaccatgatattccaatatgtgaaagttgtaagttgacagtgga 514
 QY 481 cagaactggag 491
 DB 515 cagaactcgag 525

RESULT 3
 AAV02790

ID AAV02790 standard; DNA; 649 BP.

XX
 AC AAV02790;

DT 27-APR-1998 (first entry)

XX Human partial complement factor H cDNA.

DE Complement factor H; tumour associated antigen; renal cancer;
 KW urogenital cancer; medicament; modulator; ss.

OS Homo sapiens

XX W09738136-A1.

XX 16-OCT-1997.

XX 09-APR-1997; 97WO-US05710.

XX 06-MAR-1997; 97US-0812481.

XX 09-APR-1996; 96US-0015083.

XX 09-APR-1996; 96US-0630048.

XX 06-MAR-1997; 97US-0038614.

XX (BARD-) BARD DIAGNOSTIC SCI INC.

XX
 PI Enfield DL, Hass GM, Kinders RJ;
 XX WPI; 1997-512742/47.
 DR P-PSDB; AAW39155.
 XX Treating or screening for cancer, e.g. renal or urogenital cancer -
 PT by modulating or detecting tumour associated human complement factor
 PT H related antigen, or nucleic acid encoding it
 XX Example 6B; Fig 6A; 104pp; English.
 XX This partial cDNA sequence encodes a region of the human
 CC tumour-associated complement factor H (CFH). This sequence is used
 CC in the identification of DNA encoding complement factor H related
 CC proteins and antigens from clone PRB9FH410 (see AAV02791). The
 CC detection of this protein and a CFH antigen can be used in screening or
 CC for the treatment of renal or urogenital cancer, e.g. bladder, cervical
 CC or prostate cancer. Agents that may modulate this antigen could be used
 CC in the manufacture of a medicament for the treatment of a tumour cell.
 XX
 SQ Sequence 649 BP; 218 A; 102 C; 153 G; 176 T; 0 other;

Query Match 10.1%; Score 427.2; DB 18; Length 649;
 Best Local Similarity 78.7%; Pred. No. 1.4e-91;
 Matches 510; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 369 acatgtgtagaagggtaccacactatttagtgaaattgattacogtgaattgatgcagat 428
 DB 1 acatgtaagtgggggtatcaattgtcagtgagattcaattacogtgaattgatgcagat 60
 QY 429 ggggtgaccacatgatattccaatatgtgaagttgtgaagttgttcagtgacagaactg 488
 DB 61 ggaatggacaaatgatttctcattatgtgaagttgtgaagttgttcacagtgacagacca 120
 QY 489 ggaatggagaatgtgagttgagtcgagccgacacacagaccaggaattattttggacag 548
 DB 121 gagaatggaaaaattgtcagtagtcaatggaaacacagatcgggaataccattttggacaa 180
 QY 549 gtgtgacgttttgaatgcaactccggttcagattgaagattgaagacagaaaatgcactgc 608
 DB 181 gcaatgacgttttgaatgcaactccggttcagattgaagattgaagacagaaaatgcactgc 240
 QY 609 tcataaaatggcctctggagcaatgaaagccacagttgtgaaatttctgcctgccca 668
 DB 241 tcagacgatgttttggagtaagagaacacaaagtggtggaatttcacgcaaatcc 300
 QY 669 ccacgagttgaaaaatggagatggtatatctgaaacacagtttacaaaggagaaatgaaaga 728
 DB 301 ccagatgttataaattggatctctctatctcagaagattatttataagagagaatgaacga 360
 QY 729 ttccaataaaatgtaagcaaggttttggtaaaaagaaagaggggagatgtctgtcacg 788
 DB 361 ttccaataaaatgtaagcaaggttttggtaaaaagaaagaggggagatgtctgtcacg 420
 QY 789 ggtctggatggaattcctcagccttcctgtgaaagaaatgacatgtttgactccatatatt 848
 DB 421 gaattctggatggcgtccgttgccttcattgtgaagaaaaatcatgtgataatccttatatt 480
 QY 849 ccaaatgttatctacacacctccacaggattaaacacagaattgtatgaaatcagatat 908
 DB 481 ccaaatgttgactactaccctttaaggattaaacacagaactggagatgaaatcagctac 540
 QY 909 gaatgtaaaaatggcttctatctcagcccgatcacctgttttcaaagtgtacaattact 968
 DB 541 cagtgtagaattggtttttctcctgcgaacccggggggaacacagcccaatgcagaagtact 600
 QY 969 ggctggatccctgtcctcgaagatgtagcttgaaacctgtgattttcca 1016
 DB 601 ggcgtgataacctgtccgagatgtaccttgaaacctgtgattttcca 648

```
RESULT 4
AAV02794
ID AAV02794 standard; DNA: 688 BP.
AC AAV02794;
XX
XX
XX 27-APR-1998 (first entry)
XX
DE Clone PRBS3FH2910 #3.1 CFH related protein DNA fragment.
KW Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medicament; modulator; ss.
XX
OS Synthetic.
XX
XX WO9738136-A1.
XX
XX 16-OCT-1997.
XX
XX 09-APR-1997; 97WO-US05710.
XX
XX 06-MAR-1997; 97US-0812481.
XX
XX 09-APR-1996; 96US-0015083.
XX
XX 09-APR-1996; 96US-0630048.
XX
XX 06-MAR-1997; 97US-0038614.
XX
XX (BARD-) BARD DIAGNOSTIC SCI INC.
XX
XX Enfield DL, Hass GM, Kinders RJ;
XX
XX WPI: 1997-512742/47.
XX
XX P-PSDB; AAW39158.
XX
XX
XX Treating or screening for cancer, e.g. renal or urogenital cancer -
XX by modulating or detecting tumour associated human complement Factor
XX H related antigen, or nucleic acid encoding it
XX
XX Example 6B; Fig 7A; 104pp; English.
XX
XX This partial cDNA sequence is present in clone PRBS3FH2910 #3.1 and
XX encodes a complement factor H related protein with homology to a region
XX of the human tumour-associated complement factor H (CFH). The detection
XX of this protein and a CFH antigen can be used in screening or for the
XX treatment of renal or urogenital cancer, e.g. bladder, cervical or
XX prostate cancer. Agents that may modulate this antigen could be used in
XX the manufacture of a medicament for the treatment of a tumour cell.
XX
XX Sequence 688 BP; 225 A; 129 C; 161 G; 173 T; 0 other;

Query Match 9.6%; Score 405; DB 18; Length 688;
Best Local Similarity 74.5%; Pred. No. 2.6e-86;
Matches 510; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 2885 agaaagttaccatattggagagggttacatcaattgtctgaagccttgggaattga 2944
Dn ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4 agacagtatcagtaggagaagaagttacgtacaaaattgtttgaaggttttggaaattga 63
QY 2945 tggaccagcatttataatgtgttagggagacagtgtgtgaacctcccaaatgcataaa 3004
Dn ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
64 tgggctgcgaattgcaaatgttaggagaaaaatggtctcaccctccatcatgcataaa 123
QY 3005 aactgattgtgcaacttgcacatttgaattgaaattgccaaacccagacagaaaaa 3064
Dn ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
124 aacagattgtccatttaccctcagtttgaattgaaattgccaaacccagacagaaaa 183
QY 3065 atcatcaggtcagagacaacattcagatcagatgtcaccctccgtatcgaattgagg 3124
Dn ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 ttgtataagcggttgagcaagtgactacacttgcacacatttgcacaaattgcaattg 243
QY 3125 ctctgacattgtcacatgtgtttaacgaagtgattggacagccggtatgcagaagataa 3184
Dn ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
244 ajcscgtaagtgaacatgcatttaacagatggacagagggaagggcccaacatgcagagacac 303
```

CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.
CC Agents that may modulate this antigen could be used in the manufacture of
CC a medicament for the treatment of a tumour cell.
XX
SQ Sequence 581 BP; 198 A; 85 C; 139 G; 159 T; 0 other:

Query Match 8.9%; Score 376.8; DB 18; Length 581;
Best Local Similarity 78.7%; Pred. NO. 1.2e-79;
Matches 450; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

Qy	369	acatgtgatgaagggtaccacactattagtgaaattgattacogtgaattgtgatgcagat	428
Db	1	acatgtaatgagggtgatcaattctgctagtgagatcaattaccogtgaattgacacagat	60
Qy	429	gggtggaccaatgaattccaatatgtgaagtgtgaagtgcttgccagtgcagaactg	488
Db	61	ggatggaccaatgatactcctatgtgaaagtgtgaaagtgttaccagtgacagcacca	120
Qy	489	gagaatggaaagaatttgagtggtgcagcgcgaaccagacaggaatatattttggacag	548
Db	121	gagaatggaaaaattgtcagtagtgcgaatggaaccagatcggaataccattttggacaaa	180
Qy	549	gtgtgtacgcttgaatgcaactccggtctcaagattgaaaggacagaaaagaattgcactgc	608
Db	181	gcagtacggttgtatgttaactcaggctacaagatggaaggagatgaagaaatgcattgt	240
Qy	609	tcataaaatggcctctggagcaaatgaaaagccaagtgtgtgaaattttctgcctgcca	668
Db	241	tcagacgatggttttggggtaagagaaaccaaagtgtgtgaaatttcatgcaaatcc	300
Qy	669	ccacgagttagaaatggagatgggtatatactgtaaacccagtttacaaggagagaatgaaaga	728
Db	301	ccagatgtataaatggatcctctatatctcagaagattatttataaggagagaatgaacga	360
Qy	729	ttccaatataaattgaagcaagggtttgtgtacaaagaagaagggtgctgtctgcacg	788
Db	361	tttcaatataaattgaacatgggtttatgaatacagtgaaaggagagatgctgtatgcact	420
Qy	789	ggttctggatggaatcctcagccttctgtgaaagaatgacatgtttgactccatatatt	848
Db	421	gaaatctggatggcgctccgttgccttcatgtgaaagaataatcatgtgataatccttatatt	480
Qy	849	ccaaatgggtatctacacacctcacaggtattaacacagaattgattgaatcagat	908
Db	481	ccaaatgggtgactactcacctttaaggattaaacacagaactggagatgaaataacgtac	540
Qy	909	gaatgtaaaaaatggcttcttatcctgcgaacccg	940
Db	541	cagtgtagaaatggtttttatcctgcgaacccg	572

RESULT

RESOL 0
AAV02795
ID AAV02795 standard: DNA: 590 BP.

AA
AC
AAV02795:

DT 27-APR-1998 (first entry)

DE
XX
Clone pRBS3FH2910 #4.1 CFH related protein DNA fragment.

AA
KW Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medication; modulator; SS.

OS Synthetic.

PN W09738136-A1

16-OCT-1997

AA 09-APR-1997: 97WO-US05710.

PR	06-MAR-1997;	97US-0812481.	
PR	09-APR-1996;	96US-0015083.	
PR	09-APR-1996;	96US-0630048.	
PR	06-MAR-1997;	97US-0038614.	
XX	(BARD-) BARD DIAGNOSTIC SCI INC.		
XX			
XX	Enfield DL, Hass GW, Kinders RJ;		
XX			
XX	WPI: 1997-512742/47.		
DR	P-PSDB; AAM39159.		
XX			
PT	Treating or screening for cancer, e.g. renal or urogenital cancer -		
PT	by modulating or detecting tumour associated human Complement Factor		
PT	H related antigen, or nucleic acid encoding it		
XX			
PS	Example 6B; Fig 7A; 104pp; English.		
XX			
CC	This partial cDNA sequence is present in clone PRBS3FH2910 #4.1 and		
CC	encodes a complement factor H related protein with homology to a region		
CC	of the human tumour-associated complement factor H (CFH). The detection		
CC	of this protein and a CFH antigen can be used in screening or for the		
CC	treatment of renal or urogenital cancer, e.g. bladder, cervical or		
CC	prostate cancer. Agents that may modulate the antigen could be used in		
CC	the manufacture of a medicament for the treatment of a tumour cell.		
XX			
XX	Sequence 590 BP: 198 A: 112 C: 138 G: 142 T: 0 other:		

Query Match 7.9%; Score 334.4; DB 18; Length 590;
Best Local Similarity 73.3%; Pred. No. 1.5e-69;
Matches 428; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

Qy	3021	ttgccacatttgaaattgccaaaccgacagaaaaaataatcatcaggtcagg	3080
Db	3	ttcctagcttggaaatcgctacccattggaagaaaggatttgtataaagcgggt	62
Qy	3081	gaacaagtgaacttcagatgtccacctcggatcgatggaatggatgctctgacattgtcaca	3140
Db	63	gagcccgtagcttacacttgtcaacatatcacaaaaggatggagccagtaatgaaca	122
Qy	3141	tgtgttaatacagaagtggattggacagcgggtatgcaagaataattcctgtgtgaatcca	3200
Db	123	tgcattaatagcagatggacagaaggccaacatgcagagacacctcctgtgtgaatccg	182
Qy	3201	ccacatgtgccaaatgctactatacaaaaggacaaagactaaataatccatctgtgtgac	3260
Db	183	cccacagtcacaaatgcttatatgtgcgacagatgataaataatccatctggtgag	242
Qy	3261	aaagtacggtatgactgttaataaaacctttgaattatttggggaaagtggagtgatgac	3320
Db	243	agagtacggtatcaatgtgaggccctcatgaattgttggggatggaagtgaatgtgt	302
Qy	3321	caaaacgggatttggacagaaaccccgaaatgc aaagattcaacagggaaatgtgggcct	3380
Db	303	ttaaatggaaactgacggaaaccacctcaatgcaaaattctacagaaatgtgggcc	362
Qy	3381	cctccacctattgacaattggagacataacctccttgttcattaccagtatatgcaccatta	3440
Db	363	cctccacctattgacaattgggacatacttctcctcgttgcagtatagtctccagct	422
Qy	3441	tcatcagttgaataatcaatgcagaacttattctacttaagggaataagatagtaaca	3500
Db	423	tcatcagttgagtcaccaatgccagaacttgtcataacttgagggttaacagcgaaataca	482
Qy	3501	tgtagaaatggaaaagtgttctcagccaccaacctgcttacatgcatgtgtgaaccagaa	3560
Db	483	tgtagaaatggacaatggtcagaaccacaaaagtcttacatccgtgtgtaatatcccca	542
Qy	3561	gatattatggaanaacataatatatgtctcagatggaggggaaaa	3604
Db	543	gaattatggaanaattataacataacatgaattaaaggtggacagccaa	586

RESULT 7
AAC77947
ID AAC77947 standard; cDNA; 1293 BP.
XX
AC AAC77947;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated gene sequence SEQ ID NO:341.
XX
DE Human; cancer associated gene; cancer antigen; detection; cancer;
XX diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
XX antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
XX antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
XX dermatologic; neuroprotective; thrombolytic; coagulant; nootropic;
XX vasotropic; antipsoriatic; angiogenic; gene therapy; inflammation;
XX immune disorder; haematopoietic cell disease; autoimmune disorder;
XX allergic reaction; graft versus host disease; organ rejection;
XX haemostatic; thrombolytic; cardiovascular disorder; infection;
XX neurological disease; drug screening; ss.
XX
OS Homo sapiens.
XX
XX WO200055350-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05882.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2000-587533/55.
XX P-P5DB; AAB43738.
XX
XX Novel isolated nucleic acids comprising sequences encoding peptides
XX useful for treating or diagnosing e.g. cancer -
XX
XX Claim 1; Page 892; 2352pp; English.
XX
XX AAC77607 to AAC78448 encode the human cancer associated proteins given
XX in AAB43398 to AAB44239. The proteins can have activities based on the
XX tissues and cells the genes are expressed in. Example of activities
XX include: cytostatic; proliferative; vulnery; immunomodulator;
XX antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
XX antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
XX dermatologic; neuroprotective; cardiant; thrombolytic; coagulant;
XX neotropic; vasotropic; antipsoriatic and angiogenic. The
XX polynucleotides and polypeptides can be used for preventing, treating or
XX ameliorating medical conditions and diagnosing pathological conditions.
XX polynucleotides, polypeptides, antibodies, agonists and antagonists from
XX the present invention may be used to treat immune disorders by activating
XX or inhibiting the proliferation, differentiation or mobilisation of
XX immune cells, to treat disorders of haematopoietic cells, autoimmune
XX disorders, allergic reactions, graft versus host disease and organ
XX rejection, modulate haemostatic or thrombolytic activity, modulate
XX inflammation, cancers, cardiovascular disorders, neurological disease and
XX bacterial or viral infections. The peptides, nucleotides, antibodies,
XX agonists and antagonists may be also be used in drug screens. AAC78449 to
XX AAC78457 and AAB44240 represent sequences used in the exemplification of
XX the present invention.
XX
XX Sequence 1293 BP; 427 A; 242 C; 247 G; 367 T; 10 other;

Query Match 7.7%; Score 325.2; DB 21; Length 1293;
Best Local Similarity 67.7%; Pred. No. 3e-67;
Matches 459; Conservative 8; Mismatches 206; Indels 5; Gaps 1;

QY 3174 tgcaaaagataatcctctgtgtgaatccaccacatgtgccaatgtctactatactacaagg 3233
DB 512 tccactgacactctctgtgtgaatccgccacagtaacaaatgcyatatastgtcgaga 571
QY 3234 cacaagactaaatataccatctctgtgtgacaaaagtcagcttatgactgttaataaaacctttgaa 3293
DB 572 cagatgagtaaatatccatctctgtgtgagagtcagcttatsaatgttggagcccttatgaa 631
QY 3294 ttatttgggaagtgaagtgtgccaacacgggatttgacagaaacccgaaatgc 3353
DB 632 atgtttgggaagtgaagtgtgttttaaatggaacctggagcgaacccctcaatgc 691
QY 3354 aaagattcaacagggaatgtgtgctctccacctattgacaatggagacatcaacctcc 3413
DB 692 aaagattctacrgaaaatgtgtggccctccacctattgacaatggagacatcaacctcc 751
QY 3414 ttgtcattaccagtatatgcaccattatcatcagttgaatcaatgccagaaactattat 3473
DB 752 ttcccgttgtcagtatatgctccagcttcacagttgagtaaccaatgccagaaactgtat 811
QY 3474 ctacttaagggaataagatagtaacatgataaagtagaataatggacaatggtctcagccacaacc 3533
DB 812 caacttgagggttaacaagcgaataacatgtagaataatggacaatggtctcagccacaacc 871
QY 3534 tgtttacatgcattgtgtataccagaagatattatggaaaaacataataatagttctcaga 3593
DB 872 tgtttacatgcattgtgtatataatcccgagaataatggaaaattataacatagcattagg 931
QY 3594 tggagggaataatgcaagaattatttcccaatcagggggagaatattgaattcctgtgtaa 3653
DB 932 tggacagccaaacagagagctttatttggagaaacagtgagtaacagtgagttgtgtaa 991
QY 3654 cctggatatagaataatcagaggatcacctcgttctgcacaaagtgcattgaggggtcac 3713
DB 992 cggggatatacgtcttctcagcgttctcacacattgcgaacacacatgttggatgggaaa 1051
QY 3714 atcaattatcccaactgtgtataaaatcgctatacaaatatttagtaaaccttatggatga 3773
DB 1052 ctggagtagtccaactgtgcaaaaagatagaatcaatcataaartgcacacacctttattca 1111
QY 3774 gaattgacatgtatattactaataacagtttgaatttaccatttaaatattgtttagctca 3833
DB 1112 gaactttagtattaaatcagttctyaatttcattt----ttwatgtattgtttactcc 1166
QY 3834 ttctctctaataagtata 3851
DB 1167 tttttattcactacgcaaa 1184
RESULT 8
AAC39793/C
ID AAC39793 standard; DNA; 759 BP.
XX
AC AAC39793;
XX
DT 02-JUL-1999 (first entry)
XX
DE Gastric cancer associated gene.
XX
XX Cancer associated antigen; diagnosis; research; treatment; human;
XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX prostate cancer; ss.
XX
XX Homo sapiens.
XX
XX WO9904265-A2.
XX
XX 28-JAN-1999.
XX
XX 15-JUL-1998; 98WO-US14679.
XX
XX 22-JUN-1998; 98US-0102322.
XX
XX 17-JUL-1997; 97US-0896164.

Db	4	agacagttatcagtatggagaagaagttacgtacaaaatgttttgaaggltttggaattga	63
Qy	2945	tgcagcagcaattattaaatgtgtagaggacagtgtgtcagaacctcccaaatgcataaa	3004
Db	64	tgggctgcgaatgcgaattgttaggagaaaaatgtctcacctcccatcatgcataaa	123
Qy	3005	aactgattgtgacaaccttgccacatttggaaattgcaaacccgacagaaaaagaaaaaaa	3064
Db	124	aacagattgtctcagtttaccctagcttggaaatgccatacccatgggagagaagga	183
Qy	3065	atcatacaggtcagggaacaagtgcattcagatgtccacctcgcgtatcgaaatggatgg	3124
Db	184	tttgtataaggcgggtgagcaagtgaacttacactgtgcaacattatcacaaatggatgg	243
Qy	3125	ctctgacattgtcacatgtgttaatacgaagtggattggacagocggtatgcaaaagataa	3184
Db	244	agccagtaatgtaacatgcattaatagcagatggcagaggaaggccaactcagagacac	303
Qy	3185	ttcctgtgaaatccaccacatgtgccaaatgctactatactaacaaaggccacagactaa	3244
Db	304	ctcctgtgtgaatccgcacagtagcaaaaatgcttatatagttgtagacagatlgagtaa	363
Qy	3245	atatccatctgttgacaagtagcttatgtactgttaataaaccttttgaattatttgggga	3304
Db	364	atatccatctcgttgagagatgactgttatcaaagttagggcccttatgaaaatgtttgggga	423
Qy	3305	agtgtgaagtgtatgccaaaacgggaatttgacagaaacccagaaatgcaaaagattcaac	3364
Db	424	tgaagaagtgtatgtgtttaaattggaaactgggcaggaaacctcaatgcaaaagattctac	483
Qy	3365	agggaaatgtgggcccctccaccctattgacaatggagacatcacctc	3412
Db	484	aggaaaatgtggggcccctccaccctattgacaatgggagacatctacttc	531

RESULT 10

AAV02796

ID AAV02796 standard; DNA; 472 BP.

XX

AC AAV02796;

[illegible]

DI 27-APR-1998 (first entry)
XX

Human partial complement factor

[illegible]

KW Complement factor H; tumour ass

KW urogenital cancer; medicament;

XXS

US
Homo sapiens.

XX
PN W09738136-A1

XX
. TX OCTOC 65OM NI

PD 16-OCT-1997.

XX

PF 09-APR-1997; 97WO-US05710.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

PR 08-MAR-1997; 97US-0812481.
 PP 09-APR-1996: 96US-0015083

PR 09-APR-1996: 96US-0630048

PR 06-MAR-1997; 97US-0038614.

XX

PA (BARD-) BARD DIAGNOSTIC SCI INC

This partial cDNA sequence encodes a region of the human
 tumour-associated complement factor H (CFH). This sequence is used
 in the identification of DNA encoding complement factor H related
 proteins and antigens from clone p253RH2576 (see AA02797 and AA02798).
 The detection of this protein and a CFH antigen can be used in screening
 or for the treatment of renal or urogenital cancer, e.g. bladder,
 cervical or prostate cancer. Agents that may modulate this antigen could
 be used in the manufacture of a medicament for the treatment of a tumour
 cell.

Query Match	5.8%	Score 244.8	DB 18;	Length 472;
Best Local Similarity	71.2%	Pred. No. 2.6e-48;		
Matches 349: Conservative		0: Mismatches 117;	Indels 24;	Gaps 1;

Qy	2490	attgaaccacccgtgaatactctggaatcggaagaaagatactgtctcttgcgaagatcgtt	2549
Db	7	atgacaccacactgaattatcggaatcggaagaaagtatctgtctcttgcgaagaaat	66
Qy	2550	tacctaaactcaggccagagaagaaatggtgtgtaaacatggaagggtggcagtcgttacca	2609
Db	67	tatctaattcagggaagagaagaattacatgcaaaagtggagaatggaagtcgaataacca	126
Qy	2610	cgtcgacgagaaaaattccatgttccagcccccataaaattgaacatgcatctattaag	2669
Db	127	ctctgtgtgaaaaaaattccatgttcaacaccctcagatagaacacggaacattaat	186
Qy	2670	tcgcccaggctccacagaagagagagagatttaattgagtcccagcagttatgaacacgga	2729
Db	187	tcatcagaatcttcaacaga-----aaattatcacatgag	223

RESULT 11

AA34737

ID AAX347

XX

AC AAX347

XX	XX
XX	XX

YY
06-300

Human

XX
XX
XX

KW Human

KW gene t

KW tissue

KW **secret**


```
QY 2766 atattcgaagaaatagggttaacctgcaacatgggaatggagctctctgctgtgt 2825
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 217 atctctgagaagaatgaaacaacatgctacatgggaatggagttctccacctcagtg 276
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2826 atctggaataacctgtggagcccccaccctcaattcctctctgtgtattgtttctcatgaacta 2885
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 277 gaaggctcctctgtaaatctccacctcagatttctcatgtgtgtgtgtagctcacatgtca 336
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2886 gaaatttaccatattggagagagattacatacaataattgtttctgaaggct 2934
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 337 gacagttatcagatcggagaagaattcagtcacaataattgttgaagggt 385
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
AAV02797
ID AAV02797 standard; DNA; 385 BP.
XX
AC AAV02797;
XX
DT 27-APR-1998 (first entry)
XX
DE Clone pZS3FH2576 #3 CFH related protein DNA fragment.
XX
KW Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medicament; modulator; ss.
XX
OS Synthetic.
XX
PN W09738136-A1.
XX
PD 16-OCT-1997.
XX
PF 09-APR-1997; 97WO-US05710.
XX
PR 06-MAR-1997; 97US-0812481.
XX
PR 09-APR-1996; 96US-0015083.
XX
PR 09-APR-1996; 96US-0630048.
XX
PR 06-MAR-1997; 97US-0038614.
XX
PA (BARD-) BARD DIAGNOSTIC SCI INC.
XX
PI Enfield DL, Hass GM, Kinders RJ;
XX
DR WPI; 1997-512742/47.
XX
DR P-PSDB; AAW39161.
XX
PT Treating or screening for cancer, e.g. renal or urogenital cancer -
PT by modulating or detecting tumour associated human complement Factor
PT H related antigen, or nucleic acid encoding it
XX
PS Example 6B; Fig 8A; 104pp; English.
XX
CC This partial cDNA sequence is present in clone pZS3FH2576 #3 and
CC encodes a complement factor H related protein with homology to a region
CC of the human tumour-associated complement factor H (CFH). The detection
CC of this protein and a CFH antigen can be used in screening or for the
CC treatment of renal or urogenital cancer, e.g. bladder, cervical or
CC prostate cancer. Agents that may modulate this antigen could be used in
CC the manufacture of a medicament for the treatment of a tumour cell.
XX
SQ Sequence 385 BP; 127 A; 68 C; 82 G; 108 T; 0 other;

Query Match 4.6%; Score 192.6; DB 18; Length 385;
Best Local Similarity 69.9%; Pred. No. 6e-36;
Matches 286; Conservative 0; Mismatches 99; Indels 24; Gaps 1;

QY 2526 gatatctgttttcccaagatgggttacctaactcaggcccccagagaagaatggtgtataa 2585
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 gatatctgttttcccaagaaattatctaatcaggagggggaagaataatcacatgcaaa 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2586 catggagggtgcagctgtccaccgctgcagcgaaataattccatgttcccagccccct 2645
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
DB 61 gatgaagatggcagtcactcactctgtgtgaaaaattccatgttcacaaccacct 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2646 aaaaatgaacatggatctatttaagtcgcccaggctctcagaagagagagatttaatt 2705
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 cagatagaacacggaaccattattcattccaggctctccaaga----- 164
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2706 gagtccagcagttatgaacacggaactcattcagctattgtctgtagagatggattcaag 2765
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 165 -----aattatgcacatgggactaaattgagttatactgtgaggggtgttcagg 216
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2766 atatctgaagaaataggglaacctgcaacatgggaatggagctctctgctctgtgt 2825
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 217 atatctgaagaaatgaaacaacatgctacatgggaatggagttctccacctcagtg 276
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2826 gttggaataacctgtggagcccccaccctcaattcctctctgtgtattgtttctcatgaacta 2885
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 277 gaaggctcctctgtaaatctccacctcagatttctcatgtgtgtgtgtagctcacatgtca 336
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2886 gaaatttaccatattggagagaggtttacatacaataattgtttctgaaggct 2934
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 337 gacagttatcagatcggagaagaagttacgtacaaatgttttgaagggt 385
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
AAF31736/c
ID AAF31736 standard; DNA; 137 BP.
XX
AC AAF31736;
XX
DT 10-APR-2001 (first entry)
XX
DE Novel rat gene fragment ACETA48.
XX
KW Rat; ACETA; acetaminophen; differential gene expression;
KW hepatotoxic agent; pericentral hepatic necrosis;
KW non-steroidal anti-inflammatory drug; ds.
XX
OS Rattus sp.
XX
PN W0200102609-A2.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-US40292.
XX
PR 02-JUL-1999; 99US-0142335.
XX
PR 29-JUN-2000; 2000US-0607539.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Gould-Rothberg BE, Dipippo VA, Milroy L, Daniels KK;
XX
DR WPI; 2001-138157/14.
XX
CC Screening a test agent for hepatotoxicity, useful for identifying
CC agents that damage the liver, comprises measuring differential
CC expression of one or more ACETA nucleic acids in a test cell population
XX
PS Disclosure; Page 21; 68pp; English.
XX
CC The present sequence is an ACETA nucleic acid whose expression is
CC modulated at least 8-fold in the presence of acetaminophen. The
CC differentially expressed ACETA nucleic acids may be used to identify
CC agents that damage the liver, i.e. hepatotoxic agents that cause, for
CC example, pericentral hepatic necrosis, and to identify poisoning
CC associated with ingestion of non-steroidal anti-inflammatory drugs,
CC such as acetaminophen. These nucleic acids may also be used as
CC hybridisation probes to identify ACETA-encoding nucleic acids and
CC fragments for use as polymerase chain reaction (PCR) primers for
CC amplification or mutation of ACETA nucleic acids. ACETA proteins may
CC be used as immunogens to raise anti-ACETA antibodies.
```

```

XX SQ Sequence 137 BP; 32 A; 29 C; 33 G; 43 T; 0 other;
Query Match 2.7%; Score 115.6; DB 22; Length 137;
Best Local Similarity 96.7%; Pred. No. 7.9e-18;
Matches 118; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2445 aaaaattctgcccctctcccccgcagagattccaaatgcccagtgattgaaaccaccgtg 2504
Db 122 AAAAGATTCTGCCCTCTCCCTCCACAGATTCCAAATGCCCAAGTGATTGAAACACAGTG 63
QY 2505 aaatacttgatggagaaaaatctctgttcttgcgaagatggtttacctaaactcaggac 2564
Db 62 AAATACTTGATGGAGAGAGATATCTGTCTTTGCCAAGATGGTTTACCTAACTCAGGC 3
QY 2565 cc 2566
Db 2 CC 1

RESULT 15
ID AAN82402 standard; DNA; 2177 BP.
XX AC AAN82402;
XX DT 26-NOV-1990 (first entry)
XX DE B subunit of human Factor XIII.
XX KW B sub-unit; factor XIII; blood clot stabilisation; fibrin polymer;
XX KW crosslink; scleroderma; haemorrhage; ulcerative colitis; ss.
XX OS Homo sapiens.
XX FH Key
XX FT CDS 2..1521 Location/Qualifiers
XX FT /*tag= a
XX FT /product=b subunit
XX FT misc_RNA 2..58
XX FT /*tag= b
XX FT /label=leader_sequence
XX FT CDS 59..1521
XX FT /*tag= c
XX FT /product=mature b subunit
XX PN A08778694-A.
XX PD 31-MAR-1988.
XX PF 21-SEP-1987; 87AU-0078694.
XX PR 19-SEP-1986; 86US-0909512.
XX PA (ZYMO-) ZYMOGENETICA INC.
XX PI Davie E, Seale RL, Ichinose A, Holly JA, Parker GE;
XX WPI; 1988-140637/21.
XX DR P-PSDB; AAP82921.
XX FT New DNA sequences encoding sub-units of factor 13 - and corresponding
XX FT expression vectors and transformed host cells.
XX PS Disclosure; ; p; English.
XX CC The carboxyl-terminal Thr (nucleotides 1979-1981) is followed by
XX CC a stop codon (TAG), a 187 bp noncoding sequence, and a poly(A) tail
XX CC of 9 bp. The polyadenylation or processing signal of AAAAA
XX CC was identified 19 nucleotides upstream from the poly(A) tail.
XX CC When cultured the host cells will produce the polypeptide which
XX CC can be assembled to factor 13, which stabilises blood clots by

```

```

CC crosslinking fibrin polymers. Factor 13 is useful therapeutically, e.g.
CC in cases of scleroderma, haemorrhage, ulcerative colitis etc., and
CC can be prep. in large amts. without risk of viral contamination.
CC See also AAN82401 and AAN82403.
XX SQ Sequence 2177 BP; 758 A; 358 C; 434 G; 627 T; 0 other;
Query Match 2.7%; Score 113.4; DB 9; Length 2177;
Best Local Similarity 48.8%; Pred. No. 7.7e-17;
Matches 436; Conservative 0; Mismatches 436; Indels 21; Gaps 4;

```

```

QY 1401 aatgggttttttctgaattcattatcatatgctctctaaatagaaaaaacacaggtataga 1460
Db 290 aatggttaccatctctgattgataaagtattgtataaaattcaagagaacatgcattatggt 349
QY 1461 tgtaaacaggatattgtaaaataccgcgagaaataatcaggaaataattacttctctcaa 1520
Db 350 tgcgttcagggtacaaaaccctcgaggaggaggtgaagagtggttcaatgctctctct 409
QY 1521 gatggatggtcacctcgaccctcatgattatgctcttgatatg-----cctgta 1571
Db 410 gatggatggtctctctcaaccctgttagaaaacatgaaacgtgtttggtcctctgaa 469
QY 1572 ttggagaattctgactaagaataatacacatggttttaacctcaatgacaaattagac 1631
Db 470 ttataatggaaattattccacacacagaaaacattcaagtgaaggaacaaagtacaa 529
QY 1632 tatgaatgcacattggtatgaaatgaataataataacataccaaaggctctataacatgt 1691
Db 530 tacgaatgctactggtctactacacagctgagggaagaacagagaggtagaatgt 589
QY 1692 acttatggtggtgcttagtacacctctctgttatgaaagagaatgcagcatctccctg 1751
Db 590 ctacatacggatggtctctcacacaaaatgtaccaaaattaaagtctctctttaaga 649
QY 1752 ttacaccaagattagttgttttccagagaagtaaaatacaaaagtgtgagattcgtgt 1811
Db 650 ttaattgaaatggttattttctctctgtaagcaaacctatgaagagagagatgctgt 709
QY 1812 agttctcttgcggttca--ggacacagagttggagcagatttagtcaatgctaccac 1868
Db 710 cagttttctgcatgaaaattattattcattggtgagctgtttaaattcaatgctataac 769
QY 1869 ttggtatggtcccttaatttcccaacgtgtgaagccaagtataaaatcatgtgaccaacct 1928
Db 770 ttgtgtgtaccagaatctctctgtatgcgaaggaagaagaacagatgctctctcca 829
QY 1929 cttgaaatccccgaatggggaataaaaggaacaaaaaaagtgtgaatacacagccatgtgac 1988
Db 830 cctctgccataaaactccaaaattca--aacacattcaacaacttatcgtcatggagaa 886
QY 1989 gtggtggaataatgattgcaaaccttagattctctactgaagggaacccaataaaatccagtt 2048
Db 887 atagttcatatagaattgaaacttaatttctgagatcccaatgggtcagcagaataatcgtgt 946
QY 2049 gttgacgggaagtggacaaggttgcgatatcgcttg-----agtatgagagaacatgt 2102
Db 947 gaaggtggaaaatggacagaacctccaaaatgcattgaaggacagagagaagtagcctgt 1006
QY 2103 ggagacctctctgaacttgagcatggtctgtcgaagttatctgtccctccctaccatcat 2162
Db 1007 gaggaaaccaccttcatgaaaatggtgcagcaaatttacactctaagattttatacaat 1066
QY 2163 ggagattcagtggaattcactgttacagaaaccttcacaatgattggacatcagtagtt 2222
Db 1067 ggggataaaggacatatgcatgtaaaaggcggtccctctccatggatcccaatgagata 1126
QY 2223 ttctgcattagtggaaggtggaccgagcttccctcaatgtgttgcaacagatca 2275
Db 1127 acttgtaatcgtggaaaatggacacttctcctcctgagtggttgtaaaataatga 1179

```

Search completed: August 30, 2002, 19:15:26
Job time: 11141 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 19:06:49 ; Search time 6225.59 seconds
(without alignments)
2910.950 Million cell updates/sec

Title: US-09-316-163-2
Perfect score: 866
Sequence: 1 tcagatcaactgtcccca.....ttgaattacattgaaaaa 866

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pi.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg90_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

ALIGNMENTS

RESULT 1

A91884	A91884	Sequence 2 from Patent WO9823638.	866 bp	DNA	linear	PAT 22-JAN-2000
LOCUS	A91884	Sequence 2 from Patent WO9823638.				
DEFINITION	A91884					
ACCESSION	A91884					
VERSION	A91884.1	GI:6740761				
KEYWORDS						
SOURCE		unidentified.				
ORGANISM		unclassified.				
REFERENCE		1 (bases 1 to 866)				
AUTHORS		Schwaebble, W. and Sim, R.B.				
TITLE		COMPLEMENT INHIBITOR				
JOURNAL		Patent: WO 9823638-A 2 04-JUN-1998;				
FEATURES		SCHWAEUBLE WILHELM (GB); UNIV LEICESTER (GB)				
source		Location/Qualifiers				
		1..866				
		/organism="unidentified"				
		/db_xref="taxon:32644"				
BASE COUNT		273 a 154 c 204 g 235 t				
ORIGIN						

Query Match	100.0%	Score 866;	DB 6;	Length 866;
Best Local Similarity	100.0%;	Pred. No. 1.9e-217;		
Matches 866;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy 1	tcgagtcacactgcctccagatagatccaaagacatgagagctgcagcaagaalatttggc	60		
Db 1	TCGAGTCAACTGCTCCAGATAGATCCAAAGACATGAGACTGTCAGCAACAATATTTCGGC	60		
Qy 61	ttatatatgagactgtttgtgtagcagaagattgtaaaagttctctctccagagaaaaatt	120		
Db 61	TTATATTATGACTGTTTTGTGTAGCAGAAGATTTGTAAAGTCTCTCTCCAAGAGAAAAATT	120		
Qy 121	cagaaattctctcaggttcggttcctaaacactatattcagaagcactcagcgaaacct	180		
Db 121	CAGAAATTCTCTCAGTTTCGTGGTCTGAACAACTATATTTCAGAGGCACTCAGGCAACT	180		
Qy 181	acaaatgcgcgcctggatcaccgaacacttggctactattgtaaaagtatgcagaatggag	240		
Db 181	ACAAATGCGCCCTGGATACCGAACACTTGTACTATTGTTAAAGTATGCAAGAATGCAG	240		
Qy 241	aatgggttaccttctaaccctcaaggatatagtcggaaaaagccatgtgggcaccccgag	300		
Db 241	AATGGGTACCTTCTAAACCCTCAAGAGTATGTCGGAAAAGGCCATGTGGGATCCCGAG	300		
Qy 301	acacacccctttgggtcccttaggctggcagttggatctgaaatttgaatttggtgcgaagg	360		
Db 301	ACACACCCTTTGGGTCTTTAGCTGGCAGCTTGGATCTGAATTTGAATTTGGTGCAAAAG	360		
Qy 361	tgtttatcatatgtagaagggtaccacatttagtgaaattgattaccgtagaattg	420		
Db 361	TTGTTTATPACATGTGATGAAGGGTACCAACTATTAGGTGAAATTCATTACCGTGAATGTG	420		
Qy 421	atgcagatgggtgaccaaataattccaatatgtaagttgaaagttgcctggcagtg	480		
Db 421	ATGCAGATGGGTGACCAATGATATTCCAATATGTGAAGTTGTGAAGTGTCTGCCAGTGA	480		
Qy 481	cagaactggagaaatggaaagaatttgagtcggtgcagccgaaccagaccaggaattatt	540		
Db 481	CAGAACTGGGAATGGGAAGAATTGTGACTGCTGAGCCGAACACAGCAGCAATATTATT	540		
Qy 541	ttgacaggttgtagctttgaaatgcaactccggtctcaagattgaagagcagaaagaa	600		
Db 541	TTGGACAGGTGGTACGCTTTTGAATGCAACTCCGGCTTCAAGATTTCAAGACAGAGAAAGAA	600		
Qy 601	tgcactgctcataaaatggcctctggagcaatgaaaagccacagtggtggaatttctt	660		
Db 601	TGCAGTCTCATAAATGGCCCTCTGGAGCATGAAAAGCCACAGTGTGTGCAAAATTTCCT	660		
Qy 661	gcctgcccaccacgagttgaaaatggagatggatagaaaaattcagaggatcacctccgt	720		
Db 661	GCCTGCCACCACGAGTTGAAATGGAGTGGATGATAGAAAAATTCAGAGGATCACCTCCGT	720		
Qy 721	ttcgtacaaagtgcattgagggtcacatcaattatccacttggtataaaalgcgtata	780		
Db 721	TTCTGTACAAAGTCATGTGAGGGTCACATCAATATCCACCTTGTGTATAAAATCGCTATA	780		
Qy 781	caattattagtaaaccttatggtagacactttgttagaatagcacatgatatatactaa	840		
Db 781	CAATTATTAGTAAACCTTATGGATGACACTTTGTTTAGAAATGCACATGTATATTACTAA	840		
Qy 841	tacagttgaatttacatttgaaaaa	866		
Db 841	TACAGTTTGAATTTACATTTGAAAAA	866		
RESULT 2				
LOCUS A91883				
DEFINITION A91883	Sequence 1 from Patent WO9823638.	4229 bp	DNA	linear
ACCESSION A91883				
VERSION A91883.1	GI:6740760			
KEYWORDS	unidentified.			
SOURCE				

ACCESSION AJ320522
VERSION AJ320522.1 GI:15485712
KEYWORDS complement inhibitory factor H, FH gene.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 4256)
Demberg, T., Goetze, O., and Schlaf, G.
Rat complement factor H: molecular cloning, sequencing and expression in tissues and isolated cells
unpublished
2 (bases 1 to 4256)
Demberg, T.
Direct Submission
Submitted (07-AUG-2001) Demberg T., Immunology, Georg August University of Goettingen, Kreuzbergstr 57, 37073 Goettingen, GERMANY
FEATURES
source
1..4256
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="liver"
/clone_lib="Lambda Triplex"
/dev_stage="ages 10 - 12 weeks"
/tissue_lib="CLONTECH rat liver 5' stretch plus cDNA library"
24..4209
/gene="FH"
24..3734
/gene="FH"
/codon_start=1
/product="complement inhibitory factor H"
/protein_id="CAC67513.1"
/db_xref="GI:15485713"
/translation="MRLSRIILWLTVCVAEDCKGPPPRENSIELSGWSSEQLYSE GTQATYKCRPGYRTLTGTLIVKCKNGEWPSPNSRICKRPGCFHGDTPFGSFLAVGS EFEEFAGKVVYQVYFECNQLGIDYRECDADGWINDIPICEVVKCLPVLNENGRIVSG AAEPDQYVYQVYFECNQLGIDYRECDADGWINDIPICEVVKCLPVLNENGRIVSG DGIVLPKVVYFECNQLGIDYRECDADGWINDIPICEVVKCLPVLNENGRIVSG YTPHRIKHRIIDEIYRECKNGLYPATRSPVSKITICWIPAPRSLKPCDFPQKHGR LYVESRPYFPVIGREYSYCDNGFTTPSQSDWYLRKTVNGWEPVPCRLQCFIH XVEGESYFWRRTIEGSAKQCHSYLSPNGQDTILCTENGWSPPKVRIKTCIS CDMVFENAMTKNNNTFKLNDKLDYCHIGYENYKHTGSICTYDQWSSTPSCYE RECSIPLLHQDLVFPREVYKVGDSLFSFCSRSHRGADLVQCYHFGWSNFPTECEG QVKSQDPLEIPNGEIKCTKKVESHGVDVEYDCKPFLKGNPKIQCVGDKWTLPI CVEYERTGCDLPELHSGVSKLSPVYHHGDSVEFTCTFTWIGHAVVFCISGRWTEL PQCVATDLEKCKAPKSTGIDAIHPNKNFNSVSRCHROKOEYEHISICINGRWD EPNCTRNEKRCFPPPPQIPNAQVIETTVKYLDEKSVYLCQDGYLTQGPENMCKHGR WQSLPRCTEKIPCSOPKIEHGSIKSPRSEERDLIESSEYEHGTFTSYVCDGFRIS EENRTCMNGKWSLPCRCVGPICGPPSPISPLGIVSHELESYQVGEVYNGSEGGID GPATIKCVGGOWSPPKCIKTDCLNLTFTFAKTEKKKSYRSGEQTFRCPDPYRM DGSIVTCVNTKWIQGVPCVCKNSCVNPPHVNAILRHKTKYFSGDKVRVDCNKPFE LFWGSGDVPKGTWEPKCKDSTGKCPPPPIDNGDITSLSLPVTAPLSVSYQCC NYLLKGNKIVTCRNGKWSOPPTCLHACVIPEDIMEKHNIIVLRWENAKIYSQSGENI EFMCCKPGYRKFPGSPPTKICIEGHINYPYTCV"
polyA_signal
4194..4199
/gene="FH"
polyA_site
4209
/gene="FH"
BASE COUNT 1423 a 788 c 892 g 1153 t
ORIGIN
Query Match 78.4%; Score 678.8; DB 10; Length 4256;
Best Local Similarity 99.7%; Pred. No. 3.7e-168;
Matches 680; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 10 ctgctccagatagatccaaagacatgagactgtcagcaagaattatttggtattattat 69

Db 1 CTGCTCCAGATAGATCCAAAGACATGAGACTGTGACGAGAAATATTATTTGCTTATATTAT 60
Qy 70 ggactttgttagcagaagattgtaaaggctcctcctcaagagaaaatttcagaaattc 129
Db 61 GGACTCTTTGTGTAGCAGAGATGTAAAGGCTCTCTCCAGAGAAAATTCAGAAATTC 120
Qy 130 tctcaggttcgtgctgaacaactatattcagaaggcactcaggcaacctacaaatgcc 189
Db 121 TCTCAGGTTGGTGTGAACAACTATATTAGAAGGCACTCAGGCAACCTACAAATGCC 180
Qy 190 gccctggatccagaaacacttggctactattgtaaaagatgcaagaatggagaatgggtac 249
Db 181 GCCCTGGATACCGAACACTTGGTACTATTCTAAAGTATGCAAGAAATGGAGATGGGTAC 240
Qy 250 cttctaaccatcaagatatgtcgaaagccatgtgggaatcccgagagacacacct 309
Db 241 CTTCTAACCCATCAGGATATGTGGAAAGGCCATGTGGCATCCCGGAGACACACCT 300
Qy 310 ttgggtccttttaggtggcagttggatctgaatttgaatttggcgaagggtgtttata 369
Db 301 TTGGGTCTTTAGGCTGGCAGTTGGATCTGAATTTGATTTGGTGCAGAGTTGTTTATA 360
Qy 370 catgtgatgaagggtaccactatttagtgaaattgattaccgtgaatgtgacgagatg 429
Db 361 CATGTGATGAAGGGTACCAACTCTTAGGTGAAATTTGATTTACCGTGAATGTGATGCAGATG 420
Qy 430 gdtgaccaatcatattccaatatgtgaatgtgaagtgtgagtgcttgcagtgacagaactgg 489
Db 421 GGTGGACCAATGATATTCGAATATGTGAAGTGTGTGAAGTGTGTGCGAGTGACAGAACTGG 480
Qy 490 agaatggaagaattgtgagtggtgcagccgaaccagaccaggaataattatttggscagg 549
Db 481 AGAATGGAAGAATTGTGAGTGTGCGAGCGAACCAGACAGCAAGGAATATTATTGACAGG 540
Qy 550 tggtagcgtttgaatgcaactcgcgcttcaagattgaaagacagaagaatgcaactgct 609
Db 541 TGGTAGCGTTTGAATGCAACTCCGGCTTCAAGATTGAAGGACAGAGAAATGCACTGCT 600
Qy 610 cataaatggcctctgagcaatgaaagccacagctgtgtgaaatttctgcctgctccac 669
Db 601 CAGAAAATGGCCTCTGGACCAATGAAGCCACAGCTGTGTGGAATTTCTTGCTCTCCAC 660
Qy 670 cagagttgaaatggagatgg 691
Db 661 CACGAGTTGAAAATGGAGATGG 682
RESULT 4
A91885
LOCUS A91885 Sequence 3 from Patent WO9823638. linear PAT 22-JAN-2000
DEFINITION A91885
ACCESSION A91885
VERSION A91885.1 GI:6740762
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 2715)
AUTHORS Schwaebler, W. and Sim, R. B.
TITLE COMPLEMENT INHIBITOR
JOURNAL Patent: WO 9823638-A 3 04-JUN-1998;
FEATURES SCHWAEBLE WILHELM (GB); UNIV LEICESTER (GB)
Location/Qualifiers
source 1..2715
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 863 a 518 c 618 g 716 t
ORIGIN

Matches 683; Conservative 0; Mismatches 40; Indels 0; Gaps 0;			
QY	1	tcagtgcaactgctccagatagatccaagacatgagactgtcagcaagaattatttggc	60
Db	1	TCAGTCAACTGCTCCAGATAGATCCAAGACATGAGACTGTGAGCAAGAAATTTATTGGC	60
QY	61	ttatatggagctgttctgtagcagaagattgtaaaagctctccccaagagaaaatt	120
Db	61	TTATATTATGAGACTGTTGTGTAGCAGAAGATTGTAAGGCTCTCTCCCAAGAGAAAATT	120
QY	121	cagaattctcaggttcggtctgaaacaactatattcagaagcactcaggaacact	180
Db	121	CAGAAATCTCTCAGGTTCGTGTCTGAAACAATATATTAGAAGCAGCTCAGGCAACCT	180
QY	181	acaaatgcgcgccttgataccgaacacttggtactattgtaaaagtatgcaagaatggag	240
Db	181	ACAAATGCCGCCCTGATACCGAACACTTGGTACTATTGTAAAAGTATGCAAGATGGAG	240
QY	241	aatgggtactcttaacccatacgaatgatatgtcggaaggccatgtggcatcccgagg	300
Db	241	AATGGGTACTCTTAACCCATCAAGGATATGTCGGAAGAGGCCATGTGGGCATCCCGGAG	300
QY	301	acacaccccttgggtcctttaggctggcagtgatggaattggaatttgatggatggaagg	360
Db	301	ACACACCCCTTGGGTCTTTAGGCTGGCAGTTGGATCTCAATTTGAATTTGGTGCAGAGG	360
QY	361	ttgtttatacatgtgatgaagggtaccacactattagtgaaattgataattgataattggtg	420
Db	361	TTGTTTATACATGTGATGAAGGCTACCAACTATTAGTGCAATTTGATACCGTGAATGTG	420
QY	421	atcacatgggtggacaaatgatattccaatatgtgaagtgtgaagtgttgccagtga	480
Db	421	ATCCAGATGGGTGGACCAATGATATTTCCAATATGTGAAGTTGTGAAGTGTCTGCCAGTCA	480
QY	481	cagaactgagaatgaaagaattgagtgagtgacacacacacacacacacacacacacac	540
Db	481	CAGAACTGGAGATGGAAGAAATGTGAGTGTGCACCCGAAACACAGACACAGGAATATTAT	540
QY	541	ttggacaggttggtacgctttgaatgcaactccggtctcaagattggaagcagagaaagaa	600
Db	541	TTGGACAGGTGTTACGCTTTGAATGCAACTCCGGCTTCAAGATTGAAGACAGCAAGAAA	600
QY	601	tgcactgctcataaaatggccttgagacaatgaaagccacagtgctggaattctt	660
Db	601	TGCACCTGCTCATAAATGGCTCTTGAGCAATGAAAGCCACAGTGTGTGTTGAAACCTT	660
QY	661	gctgcacacacaggttgaaatggagatggatagataagaaaattcagagatcacctccgt	720
Db	661	GTGATTTTCCACAAATCAACATGGAGCTCTGTATTATGAAGAAAGCCGGAGACCCCTACT	720
QY	721	ttc 723	
Db	721	TCC 723	
RESULT	5		
MUSH			
LOCUS			
DEFINITION	MUSH	4300 bp	mRNA
ACCESSION	Mouse Cfh locus, complement protein H gene, complete cds, clones		
VERSION	M12660		
KEYWORDS	M12660.1	GI:193724	
SOURCE	Complement protein H; protein H; serum glycoprotein. Mus musculus (strain C57/B10.WR) male liver cDNA to mRNA; Mus musculus (strain C57/B10.WR) male liver DNA; and Mus musculus (strain C57/B10.WR) male liver cDNA to mRNA.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 4300; 1 to 3425; 3474 to 4300)		
TITLE	Krisjensen, T. and Tack, B.F. Murine protein H is comprised of 20 repeating units, 61 amino acids in length		

JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967 (1986)		
MEDLINE	86233353		
COMMENT	Draft entry and clean copy sequence for [1]. [Unpublished (1986) Scripps Clinic and Res Found, La Jolla, CA 92037] kindly provided by T. Kristensen, 28-JUL-1986.		
FEATURES	Location/Qualifiers		
source	1..4300		
	/organism="Mus musculus"		
	/strain="C57/B10.WR"		
	/db_xref="taxon:10090"		
	/sex="male"		
	/tissue_type="liver"		
exon	<1..3425		
	/note="H protein"		
sig_peptide	101..154		
	/note="H protein signal peptide"		
CDS	join(101..3425,3474..3853)		
	/note="precursor"		
	/codon_start=1		
	/product="complement protein H"		
	/protein_id="AAA37759.1"		
	/db_xref="GI:387181"		
	/translation="MRLSARIILWLITVCAAECDKPPPPRENSSEILSGSWSEQLYPE GTOATV KCRPGVYRTLTGTVKVKNGKMWASNP SRICRKKPCGHPGDTFGSFLAVGS OFEGKAVVYTCDDGYQLLEIDYRECCAGDWINDIPLCEVVKCLPVTLENGRIVSG AAETQOEYFGQVAFECNSGFKIEGKHEHCSENGLSNEKPRCIVEICTPPRVENG DQINVKPVYKENERHYKCKHGVKPERGDAVCTGSGWSSQPFCEKRSPPYILNGI YTPHRIIHRSDDEIRYENGYEYPTVTSKCTPTGWIPTPCTLKPCFEPQFKYGR LYEBSLRNFPVSNKYSYKNDGFPSPGSDYDYLACTAGWSPVPCVKCVFHV VYNGDSAYWEKVYVQGSGLKQCYNGYSLONGQDTMTCTENGWSPPPKCIRIKTCSA SDIHNDGFLSESSSYALNRETSYRCKQGYVTNTEIGESITCLONGWSPSCILKS CDMPVENSITKNTKTPKLNKLDYECLVGFENYKHTKGSITCTYTGWSDTPSCYE RCSVPTLDRKLVSPRAEYKRGDLLEFSCSHRGVSDSVOCYHGWSPGPTCKG QVASCAPPLEILNGEINGAKVEYSHGEVYKDYDKPFLKGNPKIQCVYDGNMTLLP CTEERTCGDIPLEHSGSAKSVPPYHHGDSVEFICEENFTMIGHSVSCISGKWTOL PKCATDOLKCRVLKSTGIEAIKPKLTFTNTSTMDYKCRQKQYERISICINGKWD BPNCTSKTSCPPPOIPNTQVLTETVKYLDGKLSVLCDONYLTDSEEMVCKDGRMQ SLPRCLIEKIPCSOPTIEHGSINLPRSEERRDSIESSEHNGTTFTSYVCDGDFRIPE ENRITCYMGKWSPTPRCVGLPCGPPSPILGTVSLELESYQHGEVYTHGEEVYTGSGIDG PFIITEGGKNSDPKCIKTDGCVLPTVKNATIRGKSKSYRTGEQVTPRCSPTOMN GSDTITCVNSRWIGQVCKDNDSCVDPHPVNATIVTRTKNRYLHGDVREYCNKPLEL FQGVEMCENGITWKEKPSBPPTCLHACVIPENIMESHNIILKWRHTEKIYSHSGEDIEF YLLKKKKTICTNGKWSBPPTCLHACVIPENIMESHNIILKWRHTEKIYSHSGEDIEF CCKYGYKAROSPPTKINGTINPTCV"		
mat_peptide	join(155..3425,3474..3850)		
	/note="H protein"		
intron	/product="complement protein H"		
	3426..3473		
exon	/note="H cds intron"		
	3474..4300		
	/number=2		
BASE COUNT	1401 a 814 c 902 g 1183 t		
ORIGIN	554 bp upstream of XhoI site; chromosome 1.		
Query Match	68.9%; Score 596.6; DB 10; Length 4300;		
Best Local Similarity	91.5%; Pred. No. 1.7e-146;		
Matches 632; Conservative 0; Mismatches 59; Indels 0; Gaps 0;			
QY	1	tcagtgcaactgctccagatagatccaagacatgagactgtcagcaagaattatttggc	60
Db	69	TGAGTCACTTGGTCCCAAGAGATCAATATTATGAGACTGTGAGCAAGAAATTTATTGGC	128
QY	61	ttatatggagctgttctgtagcagaagattgtaaaagctctccccaagagaaaaatt	120
Ddb	129	TTATATTATGAGACTGTTGTGTCAGCAGCAGAAATTTGAAGGTCTCTCTCCCAAGAGAAAATT	188
QY	121	cagaattctcaggttcggtctgaaacaactatattcagaagcactcaggaacact	180
Db	189	CAGAAATCTCTCAGGTTCGTGTGTCAGCAGCAGAACTATATCCAGAGGACCCAGGAACT	248

```

Qy 181 acaaatgcgccttgataccgaacacttggtactattgtataaagtatcgaagaatggag 240
Db 249 ACAAAATGCGCCTGGATACCAACACTTGGCACATATTGTAAGATATCAAGAATGAA 308
Qy 241 aatgggtaccttcaaccatcaagatatgtcgaagaagccatgtggcatcccgag 300
Db 309 AATGGGTGGGCTTAACCCATCCAGGATATGTCGGAAGAAAGCCTTGTGGCATCCCGGAG 368
Qy 301 acacacccttgggtccctttaggtggcagttggatctggaatttgaatttgggtgcaagg 360
Db 369 ACACACCCCTTTGGGTCCCTTAGGCTGGCAGTTGGATCTCAATTTGAGTTGGTGCAAGG 428
Qy 361 ttgtttatcatgtgatgaagggtaccactattagtggaattggaattgattaccgtgaatgtg 420
Db 429 TGTGTTATACCTTGATGATGGGTATCAACTATTAGGTGAAATGATTACCGCTGAATGTG 488
Qy 421 atgcagatgggtggaccatgatattccaatatgtggaagtgtggaagtgtggtccagtgga 480
Db 489 GTGCAGATGGCTGATCAATGATATCCACTATGTAAGTTGTGAAGTGTCTACCTGTGA 548
Qy 481 cagaactggagaatggaagaattgtgagtggtgcagccgaaccagaccaggaataattatt 540
Db 549 CAGAACTCCGAATGGAAGAATTGTGAGTGTGTCAGCAGAAACAGACCAGGAATACTATT 608
Qy 541 ttgacaggtggtcgtcttgaatgcaactcgcgtctcaagattgaaagacagaagaaga 600
Db 609 TTGGACAGGTGGTGGCTTTGAATGCAATTTCAGGCTTCAAGATTGAAGGACATAAGGAAA 668
Qy 601 tgcactgtctataaatggcctctgggaatgaaagccacagtggtggaattttctt 660
Db 669 TTCATTGCTCAGAAATGGCCTTTGGAGCAATGAAAGCCACCGATGTGTGGAATTCCT 728
Qy 661 gcttgcacacacaggtggaataatggagatgg 691
Db 729 GCACACCACCGGAGTGGAATAAGAGATGG 759

RESULT 6
HSHT HSH 1658 bp mRNA linear PRI 11-MAR-2001
LOCUS Human mRNA for truncated form of complement factor H.
DEFINITION X07523
ACCESSION X07523.1 GI:32492
VERSION complement factor H.
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1658)
AUTHORS Ripoché,J., Day,A.J., Harris,T.J. and Sim,R.B.
TITLE The complete amino acid sequence of human complement factor H
JOURNAL Biochem. J. 249 (2), 593-602 (1988)
MEDLINE 88134059
REFERENCE 2 (bases 1 to 1658)
AUTHORS Day,A.J.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-1988) Day A.J., Dept. of Biochemistry, University
of Oxford, MRC Immunochimistry Unit, South Parks Road, Oxford OX1
3QU
COMMENT Data kindly reviewed (06-JUN-1988) by Day A.J.
FEATURES
source
1..1658
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="B-38-1 tissue-liver"
/clone_lib="pat153/pvuII/8"
<1..73
/feature="5'-UT region"
misc_feature
1..1408
misc_feature
74..127
/feature="seq. identical to full-length factor H"
sig_peptide
74..127
/feature="put. signal peptide"
74..1423
CDS

```

```

/codon_start=1
/product="complement factor H"
/protein_id="CAA30403.1"
/db_xref="GI:758073"
/db_xref="SPTREMBL:Q14570"
/translation="MRLAKIICLMLAICVAEDCNELPPRRNTEILTGSNSDQTYE
GTQAIYKCRPGYSLGNVIMVCRKGEWVNLRLKQKRCQPCGHFQDTPFGFTLTGGN
VFEGYKAVYTCNEGQQLLNGINRECDTNDIPICEVVKLPVTAENKGVSS
AMEPDREYHFGOAVRFVNSGYKIEGDEHMSDDGPMWPKPCVEISCKDNPYIPNGD
SPIQKIIYKENERFOYKCNMGVEYSEGDVCTESGWRPLPSCCEKSCDNPYIPNGD
XPLRLKHRTGDEITYOCRNFGYPATRGNTAKCTSTGWIIPAPRCTLKPDPYDKHGG
LYHNMRRPFYPAVAGKITYSTYDEHFTPSGSWDHIHCTODGWSVAPCLRCYFP
YLENGYNONGRKEFVGQKSIDVACHPGYALPKAQTTVTCMENGWSPTPRCIRVSFTL
128..1420
/mat_peptide
128..1420
/misc_feature
1409..1420
/feature="unique coding seq. not found in full-length factor
H"
1421..1423
/feature="stop codon"
1424..1658
/feature="3'-UT region"
1634..1639
/misc_feature
/feature="pot. polyA signal"
BASE COUNT 513 a 296 c 383 g 466 t
ORIGIN
Query Match 51.6%; Score 446.8; DB 9; Length 1658;
Best Local Similarity 78.4%; Pred. No. 4.6e-107;
Matches 535; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
Qy 5 g t c a a c t g t c c c a g a t a g a c a t g a g a c t g c a g a a g a a t t a t t g c t t a t 64
Db 46 G T A A A T G T C C T T T A A A A G A T C C A A A A A T G A G A C T T C T A G C A A A G A T A T T T G C C T A T 105
Qy 65 a t t a t g g a c t g t t g t g t a g c a g a a g a t t g t a a a g g t c c t c c t c c a a g a g a a a t t c a g a 124
Db 106 G T T A T G G G C T A T T T G T G T A G C A G A A G A T T G C A A T G A A C T T C C C A A A G A A T A T T T G C C T A T 105
Qy 125 a a t c t c t c a g g t c g t g t c t g a c a a c t a t t c a a a g g c a c t c a g g a c c a g g a a c c a a a 184
Db 166 A A T T C T G A C A G G T T C C T G G T C T G A C C A A A C A T A T C C A A A G C C C A G G C T A T A T A A 225
Qy 185 a t g c g c c t g g a t a c c a c a c a c t t g g t a c t a t t g t a a a g t a t c a a g a a t g g a a g a t g 244
Db 226 A T G C G C C C T G A T A T A G A T C T T T G G A A A T G T A A T A T G T A T C A G A G A G G A G A A T G 285
Qy 245 g g t a c c t t c t a c c a t c a a g g a t a t g t c g g a a a g g c c a t g t g g g c a t c c c g g a g a c a c 304
Db 286 G G T T G C T C T T A A T C C A T T A A G G A A A T G T C A G A A A A G G C C C T G T G G A C A T C C T G G A G A T A C 345
Qy 305 a c c c t t t g g g t c c t t t a g g c t g g c a g t t g g a t g a a t t g a a t t g g t g c a a a g g t t g t 364
Db 346 T C C T T T T G G T A C T T T T A C C C T T A C A G G A G A A A T G T G T T G A A T A T G T G T A A A A G C T G T 405
Qy 365 t t a c a t a t g t a g a a g g t a c c a a c t a t t a g g t g a a a t t a t t a c c t g a a t g a t g a t g c 424
Db 406 G T A T A C A T G T A A T C A G G G T A T C A A T T C C T A G G T G A G A T T A A T T A C C G T G A T G T G A C A C 465
Qy 425 a g a t g g g t g a c c a a t g a t a t t c c a a t a t g t g a a g t t g t g a a g t g c t t g c c a g t g a c a g a 484
Db 466 A G A T G G A T G G A C C A A T G A T T C C T A T A T G T G A A G T T G T G A A G T T T A C C A G T G A C A C G 525
Qy 485 a c t g a a a a t g g a a a a t t g a g t g g t g a g c c g a c c a c c a g a c c a g a a t a t t a t t t t g g 544
Db 526 A C C A G A A T G G A A A A A T T G C A G T A G T G C A A T G G A A C C A G A T C G G A A T A C C A A T T T T T G G 585
Qy 545 a c a g g t g t a c g c t t t g a a t g a a c t c c g c t c a a g a t t g a a g a c a g a a a a a a a t g c a 604
Db 586 A C A A G C A G T A C G G T T T G T A T G T A A C T C A G G C T A C A A G A T T G A A G A G A T G A A A A T G C A 645
Qy 605 c t g c t a a a a t g g c c t c t g g a g c a a t g a a a g c c a g t g t g t g g a a a t t t t t c t c c t 664

```

Db 646 TTGTTTCAGACGATGCTTTTGGAGTAAGAGAAACCAAGTGTGTGCAAAATTTTCATGCCAA 705
QY 665 gccaccacgagtgtaaaatgga 686
Db 706 ATCCCCAGATGTTATAAATGGA 727

RESULT 7
HSH 3926 bp mRNA linear PRI 08-APR-1997
LOCUS Human mRNA for complement factor H.
DEFINITION Y00716 M32093 x07525
ACCESSION Y00716.1 GI:31964
VERSION complement factor H; glycoprotein; regulatory protein.
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3926)
Day/A.J.
Direct Submission
Submitted (25-FEB-1988) Day A.J., MRC Immunochimistry Unit, Dept.
of Biochemistry, University of Oxford, South Parks Road, Oxford,
OX1 3QU
2 (bases 1 to 3926)
Ripoche,J., Day,A.J., Harris,T.J. and Sim,R.B.
The complete amino acid sequence of human complement factor H
JOURNAL Biochem. J. 249 (2), 593-602 (1988)
MEDLINE 88134059
COMMENT see Immunogenetics 27:211-214(1988) concerning TYR <-> HIS at AA
384
see X07523 for truncated (1.8 kb) mRNA sequence.

FEATURES
Location/Qualifiers
1..3926
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Bg-38-1, R2a and PE3"
/tissue_type="liver"
/clone_lib="human liver cDNA in pat153/PvuII/8"
74..127
74..3769
/codon_start=1
/product="factor H"
/protein_id="CAA68704.1"
/db_xref="GI:31965"
/translation="MRLAKIICLMLWAICVAEDCNELPPRRNTEILTGWSDDOTYPE
GTOATYKCRPGYRLGNVIMYCRKGWALNPLRCKOKRPGHGDTPRGFTLTGNN
VFEGVKAAYTCNEGYQLGELNINRECDTGDWTDIPICEVVKCLPVTAPENGLIYSS
AMEPDEVHFGQAVFVVCNSGYKLEGDEMHCSDGFWSEKPKCVCESKSPDVING
YSLPKIKRTGDEITYQCRNGFYFATRGNTAKCTSTGWIPAPRCTLKPCDYPDIKHGG
LYHENMRPFPFVAVGKYYSYCDHEFEFTPSGYNDHIHCTQDGSAPVAPGLRKYFP
SYLNGNONGKRVQGSIDVACHPGYALPKAQTVTTCMENGWSPTPRCIRVKTCSK
SSIDIENGFISESYOTALKEKAKYQCKLVYVADGETSGSIRGCKDGSAAQPTCKS
CDIPVEMNARTKNDFTWPKLNDLDEYCHDGYESNTGSTGTSIVCGYNGWSDLPICYE
RECELPKTDVHLVDPKRDQYKVEYLKFSCKPFTIVGPNVQCIHFGSLPDLPICK
EQVQSCGPPPELLNGNVKKEKTEBEYHSEVVEYCYNPRFLMKPNKIOCVDGEWTLTP
YIYEESTCGDPIPELHGWAQLSSPPYYGDSVEFNCSSEFTMIGHSITICIHGWTO
LPQVADLKLKSSNLIIIEHLKKNKEFDHNSNIRYRCRGEGWHTTVCINGRWD
PEVNSCAIQLCPPPPQIPNSHNMTTLNRYDGEKVSVLCOENYLLQEGREITCKDG
RWQSIPLCVERKIPCSOPPOIEHGTINSSRSOESYAHGTKLSYCEGFRFISENETT
CYMKWSSPQCEGLPKCPPEISHGVVHMSDSYQEGEEVYKCFEGFIDGPAIAK
CLGKNWHPSPCITCDLSLPSFENAIIPMGCKDVIYKAGEQVYTCATYKMDGASNY
TCINRWTRPTCRDTSVNPPTVONAYIVSRQMSKYPSPGERVYQCRSKYEMFGDEE
VMLGNWTEPPQCKDCKGPPPDINGDITSFPLSVAPASSVEYQCNLYQLEG
NKRITCRNGWSEPPKCLHPCVIREIMENYINIALRWTAQKLYSRITGESVEFYCKRG
YRLSSRSHTLRLTCWCKLEYPTCAK"

128..3766
/product="factor H"
994
/note="polymorphism"

mat_peptide
variation

variation
misc_feature
variation

BASE COUNT 1306 a 689 c 855 g 1076 t
ORIGIN

Query Match 51.6%; Score 446.8; DB 9; Length 3926;
Best Local Similarity 78.4%; Pred. No. 4.6e-107;
Matches 535; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 5 gtaaatgtctccagatagatccaaagacatgacagctgtcagcaagaattatttggccttat 64
Db 46 GTAAATGTCTCTTAAAGATCCAAAAATGACACTTCTAGCAAGATTTATTTCCTTAT 105
QY 65 attatggactgttctgtgtagcagaagattgtaaaggtctctccccaagagaataatcaga 124
Db 106 GTTATGGGTATTGTGTAGCAGAGATTGCAATGAACCTTCTCCAAGAGAAATACAGA 165
QY 125 aattctctcaggttctgtgctgaacaactatattcagaagggcactcaggaacacclacaa 184
Db 166 AATTCTGACAGGTTCTCTGCTGACCAACATATCCAGAGAGGCCACCCAGGCTATCTATAA 225
QY 185 atgccccttgatgataccgacacttggctactattgttaaagtatgcaaatggagaatg 244
Db 226 ATGCCGCCCTGGATATAGATCTCTTGGAAATGTAATAGTGTATGCAGGAAGGAGAAATG 285
QY 245 ggtacaccttaaccatcaagatatgtcgaaaagggccatgtgggcatcccgagagacac 304
Db 286 GGTGTGCTCTTAATCCCAATAGGAATGTGAGAAAGGCCCTGTGGACATCTCTGGAGATAC 345
QY 305 acccttgggtcccttagctggcagttggtctgattggaatttgggtgcaaaaggtgtg 364
Db 346 TCCTTTTGGTACTTTTACCCCTTACAGGAGAAATGTGTTGAATATGTTGTAAGAACTGT 405
QY 365 ttatacatgtatgaagggtaccacactattagtggaattgattaccgtgaattgtatgc 424
Db 406 GTATACATGTAATGAGGGGTATCAATGTCTAGTGAGATTAATTACCGTGAATGTGACAC 465
QY 425 agatgggtgacaaatgatattcccaatatgtgaagtgtgaagtgtgtgccaatgacaga 484
Db 466 AGATGATGGACCAATGATATTCCTTATATGTGAAGTTGTGAAGTGTGTACCAAGTACAGC 525
QY 485 actggagaatggaagaattgtgagtggtgcagccgacaccagaccaggaataattatttg 544
Db 526 ACCAGAGAATGAAAAAATTTGTGAGTAGTCAATGGAACCAAGATCGGAATACCATTTTG 585
QY 545 acaggtgtgacgtttgaatgaactccggttcgaagtgaaggtgaaggtgaaggtgaaggt 604
Db 586 ACAAGCAGTACGGTTGTGTATGTAACCTCAGGCTACAGAGATTGAAGGAGATGAAGAAATGCA 645
QY 605 ctgctcataaaatggcctctgagcaatgaaagccacagatgtgtgaaattcttgcct 664
Db 646 TTGTTTCAGACGATGTTTGTGGTAAAGAGAAACCAAGTGTGTGGAATTTTCATGCCAA 705
QY 665 gccaccacgagtgtaaaatgga 686
Db 706 ATCCCCAGATGTTATAAATGGA 727

RESULT 8
A91886
LOCUS A91886
DEFINITION Sequence 4 from Patent WO9823638.
ACCESSION A91886
VERSION A91886.1 GI:6740763

```
KEYWORDS      unidentified.
SOURCE         unidentified.
ORGANISM       unclassified.
REFERENCE      1 (bases 1 to 1532)
AUTHORS        Schwaible,W. and Sim,R.B.
TITLE          COMPLEMENT INHIBITOR
JOURNAL        Patent: WO 9823638-A 4 04-JUN-1998;
               SCHWAEIBLE WILHELM (GB); UNIV LEICESTER (GB)
FEATURES       Location/Qualifiers
               1..1532
               /organism="unidentified"
               /db_xref="taxon:32644"
BASE COUNT    523 a 275 c 290 g 444 t
ORIGIN
1 tcgagtcactgtcccgatagatccagacatgagactgtcagcaagaattatttggc 60
   |||||
Db 1 TCGAGTCAACTGCTCCAGATAGATCCAGACATGAGACTGTGACGAGAAATTATTGCG 60
   |||||
Qy 61 ttatatattgactgttgttagcagaagattgtaaggctctctcccaagagaaaatt 120
   |||||
Db 61 TTATATTATGGACTGTTTGTGTAGCAGAAGATTGTAAGGCTCTCTCCCAAGAGAAATT 120
   |||||
Qy 121 cagaattctctcaggttcgtctgaacaactatattcagaagcactcaggcaacct 180
   |||||
Db 121 CAGAAATTCTCAGGTCGTGCTGTGAACAACATATATTCAGAAGGCATCAGGCAACCT 180
   |||||
Qy 181 acaaatgcgccttgataccgaacacttggtactattgttaaaagtatgcaagaatggag 240
   |||||
Db 181 ACAAAATGCGCCTCGATACCAACACATTGGTACTATTGTAAGACTATGCAAGATGGAG 240
   |||||
Qy 241 aatgggtacctcttaaccatacaagatattgcgaaaggccatgtgggcatcccgag 300
   |||||
Db 241 AATGGGTACCTTCTTAACCCATCAAGGATATGTCGAAAAGGCCATGTGGCATCCGGAG 300
   |||||
Qy 301 acacacccttgggtcccttaggtcggcagttggtatcgaattgaattggtgcaagg 360
   |||||
Db 301 ACACACCTTTGGGTCCCTTAGGCTGGCAGTTGGATCTGAATTTGAATTTGGTGCAGAGG 360
   |||||
Qy 361 ttgtttacatgtgatgaagggtaccactattagggtgaattgattaccgta 415
   |||||
Db 361 TTGTTTATACATGTGATGAAGGGTACCAACTATTAGGTGAAATTCATTACCGTTA 415
   |||||

RESULT        9
LOCUS         HSH38
DEFINITION    Human mRNA for complement factor H 38-kDa N-terminal fragment.
ACCESSION     X04697
VERSION       X04697.1 GI:31991
KEYWORDS      complement factor H.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1427)
AUTHORS        Schulz,T.F., Schwaible,W., Stanley,K.K., Weiss,E. and Dierich,M.P.
TITLE          Human complement factor H: Isolation of cDNA clones and partial
               cDNA sequence of the 38-kDa tryptic fragment containing the binding
               site for C3b
JOURNAL        Eur. J. Immunol. 16 (11), 1351-1355 (1986)
MEDLINE       87054207
COMMENT        Clone H-19 codes for the first 108 AA of the 142-kDa fragment
               in addition to the 289 AA derived from the 38-kDa fragment.
               An additional A residue at pos. 930 not observed in a recently
               published cDNA clone caused a shift in reading frame. It represents
               not a sequencing artefact but is probably due to an error that
```

occurred during reverse transcription. Data kindly reviewed

```
(03-AUG-1987) by Schulz T.F.
Location/Qualifiers
1..1427
/organism="Homo sapiens"
/db_xref="taxon:9606"
misc_feature   1..867
               /note="38-kDa C-terminal tryptic H fragment"
CDS            join(<1..929,931..1195)
               /note="ORF"
               /codon_start=1
               /protein_id="CAB41739.1"
               /db_xref="GI:4725976"
               /translation="ILGYRSLGNVIMVCKGEWALNPLRKQKRCRPGHDPGTPTGTF
               TLTGNNVEYGVKAVYTCNEGYVOLLGEINRECDTGDNDIPICEVVKCLPVTAPEN
               GKIVSSAMEPDREYHFGOAVFCVNSGYKIEGDEEMHCSDDGFWKEPKCKVEISCKS
               PDVINGSPIKOKIIVKENERFQYKCMGYEYSEYSDGDAVCTESGWRPLPSCREKSCDNP
               YIPNGDYLPHRIKRTGDEITTCRNGEYTPATGNKTACTSTGKIPAPRCTLKPQDYP
               DIKGGLYLHENMRRFPYFVAVGKYYSYCDHEFTFPGSYMDHIHCTODGWSPAVPC
               LKCYFPYILENGYNQYGRKFVQGSIDVACHFGYALPKRAQTIVTTCMENGSPTPRCIR
               VSFTL"
misc_feature   868..1192
               /note="142-kDa N-terminal fragment"
misc_feature   930
               /note="additional A residue, causing a shift in reading
               frame"
polyA_signal   1407..1412
polyA_site     1427
BASE COUNT    435 a 254 c 333 g 405 t
ORIGIN
Query Match      37.1%; Score 321.6; DB 9; Length 1427;
Best Local Similarity 78.0%; Pred No. 4,3e-74;
Matches 387; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
Qy 191 cctgtgataccgaacacttggtactattgtataaagtatgcaagaatggagaatgggtacc 250
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 3 CCGTGATATAGATCTCTTGGAAATGTAATAATGTTATGTCAGGAAGGAGAAATGGCTTGC 62
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Qy 251 ttctaaccatcaagatatgtcggaaaggccatgtggggaatcccgaggagacacacctt 310
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 63 TCTTAATCCATTAAGGAAATGTCAGAAAAGGCCCTGTGGACATCTCTGGAGATACTCCTTT 122
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Qy 311 tgggtcccttaggtcggcagttggatctgaattgaaattggtgcaaaaggttgtttatcac 370
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 123 TGGTACTTTTACCCCTACAGGAGGAAATGTGTTGAATATGGTGTAAAAGCTGTGTATAC 182
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Qy 371 atgtgatgaagggtaccactattagggtgaattgattaccgtgaatgtgatgcagatgg 430
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 183 ATGTAATGAGGGGTATCAATTCCTAGCTCAGATTAAATACCCGTGAATGTGACACAGATGG 242
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Qy 431 gtggaccaatgatattccaataatgaaattgtgaagtgtggaagtcttgcagtagacagaactgga 490
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 243 ATGGACCAATGATATTCCTATATGTGAAGTTCTGTGAAGTGTGTACCAAGTACCAACACG 302
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Qy 491 gaatggaagaattgtgagtggtgcagccgaaccacacagcaggaattatttttggacagat 550
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 303 GAATGGAATAATTTGTCAGTAGTGCAATGGAACCATCGGGAATACCATTTTGGACAACG 362
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Qy 551 ggtacgctttgaatcgaactccggttcaagattgaaaggacagaagaataatcactgctc 610
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 363 AGTACGGTTGTGTATCTAATCTCAGGCTACAAGATTCAAGAGATGAAGAAATGCATTGTTC 422
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Qy 611 ataaaatgcctctgagcaaatgaaaagccacagtggtggaattttcttgcctgccacc 670
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 423 AGACATGGTTTTGGAGTAAAGAGAAAACCAAGTGTGTGGAAATTTTCATGCAAAATCCCC 482
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Qy 671 acgagttgaaatgga 686
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 483 AGATGTTATAATGGA 498
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
```


* 76781 79503: contig of 2723 bp in length
* 79504 79603: gap of unknown length
* 79604 82811: contig of 3208 bp in length
* 82812 82911: gap of unknown length
* 82912 85486: contig of 2575 bp in length
* 85487 85586: gap of unknown length
* 85587 88158: contig of 2572 bp in length
* 88159 88258: gap of unknown length
* 88259 90363: contig of 2105 bp in length
* 90364 90463: gap of unknown length
* 90464 93178: contig of 2715 bp in length
* 93179 93278: gap of unknown length
* 93279 95124: contig of 1846 bp in length
* 95125 95224: gap of unknown length
* 95225 98293: contig of 3069 bp in length
* 98294 98393: gap of unknown length
* 98394 101043: contig of 2650 bp in length
* 101044 101143: gap of unknown length
* 101144 104079: contig of 2936 bp in length
* 104080 104179: gap of unknown length
* 104180 105994: contig of 1815 bp in length
* 105995 106094: gap of unknown length
* 106095 108281: contig of 2187 bp in length
* 108282 108381: gap of unknown length
* 108382 110825: contig of 2444 bp in length
* 110826 110925: gap of unknown length
* 110926 113171: contig of 2246 bp in length
* 113172 113271: gap of unknown length
* 113272 115358: contig of 2087 bp in length
* 115359 115458: gap of unknown length
* 115459 117605: contig of 2147 bp in length
* 117606 117706: gap of unknown length
* 117707 118976: contig of 1271 bp in length
* 118977 119076: gap of unknown length
* 119077 121384: contig of 2308 bp in length
* 121385 121484: gap of unknown length
* 121485 123529: contig of 2045 bp in length
* 123530 123629: gap of unknown length
* 123630 125180: contig of 1551 bp in length
* 125181 125280: gap of unknown length
* 125281 126315: contig of 1035 bp in length
* 126316 126415: gap of unknown length
* 126416 127699: contig of 1284 bp in length
* 127700 127799: gap of unknown length
* 127800 129330: contig of 1531 bp in length
* 129331 129430: gap of unknown length
* 129431 130464: contig of 1034 bp in length
* 130465 130564: gap of unknown length
* 130565 132015: contig of 1451 bp in length
* 132016 132115: gap of unknown length
* 132116 134216: contig of 2101 bp in length
* 134217 134317: gap of unknown length
* 134317 135628: contig of 1312 bp in length
* 135629 135728: gap of unknown length
* 135729 137629: contig of 1901 bp in length
* 137630 137729: gap of unknown length
* 137730 139483: contig of 1754 bp in length
* 139484 139583: gap of unknown length
* 139584 141235: contig of 1652 bp in length
* 141236 141335: gap of unknown length
* 141336 142910: contig of 1575 bp in length
* 142911 143010: gap of unknown length
* 143011 144249: contig of 1239 bp in length
* 144250 144349: gap of unknown length
* 144350 145400: contig of 1051 bp in length
* 145401 145500: gap of unknown length
* 145501 147005: contig of 1505 bp in length
* 147006 147105: gap of unknown length
* 147106 148396: contig of 1291 bp in length
* 148397 148497: gap of unknown length
* 148497 150032: contig of 1536 bp in length
* 150033 150132: gap of unknown length
* 150133 151263: contig of 1131 bp in length.

FEATURES
source
Location/Qualifiers
1. .151263
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-76J6"
BASE COUNT 46018 a 28440 c 27866 g 43718 t 5221 others
ORIGIN
Query Match 15.4%; Score 133.4; DB 2; Length 151263;
Best Local Similarity 87.7%; Pred. No. 1.6e-24;
Matches 157; Conservative 0; Mismatches 21; Indels 1; Gaps 1;
Qy 689 tggatagaaaaatcgagagatcacct-cogtttcgtacaaaagtgcattgagggtcac 747
|||||
Db 34914 TGGATATAGAAAATTAAGAGGATCACCTCCCATTTGCTGTTCAAAATGCATTGATGTCACA 34855
Qy 748 tcaattatccactgtgtataaaatcgctatacaattattagtaaaccttatggatgac 807
|||||
Db 34854 TCAATTATCCCACTGGTCTGTAAATACAAATGATTTATTAGTAAACCTTTATGATGAA 34795
Qy 808 acttggttagaatgcacatgtatattactaataacagtttgaaatttgcatttgcataaa 866
|||||
Db 34794 CCTTGTATTAGAAATACATGCTTATTACTAATTCATTAATTTACATTTGAAATA 34736
RESULT 13
HSJ177P10/c
LOCUS HSJ177P10 150626 bp DNA linear PRI 26-JAN-2000
DEFINITION Human DNA sequence from clone RPI-177P10 on chromosome 1q31.2-32.1
Contains 2 isoforms of the HFI (H factor 1 (complement)) gene,
ESTs, STSS and GSSs, complete sequence.
ACCESSION AL049744
VERSION AL049744.8 GI:6782357
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150626)
AUTHORS Bird,C.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Jan 27, 2000 this sequence version replaced gi:5531531.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 1, constructed by the Sanger Centre Chromosome 1
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Emi., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WormPeP; Information
on the WormPeP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep RPI-177P10 is
from the library RPI-1 constructed at the Roswell Park Cancer
Institute by the group of Pieter de Jong. For further details see
http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RPI-177P10 The true
right end of clone RPI-15D12 is at 48650 in this sequence.
Location/Qualifiers


```

source
1. .150626
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q31.2-32.1"
/clone="Rpl-177P10"
/clone_lib="RPC1-1"
1841. .12140
repeat_region
/feature="AluSg1 repeat: matches 1. .293 of consensus"
3949. .4135
repeat_region
/feature="L1M4/D repeat: matches 5347. .5548 of consensus"
5327. .5740
misc_feature
/feature="match: GSS: Em:AQ475732"
7349. .7430
repeat_region
/feature="41 copies 2 mer aa 63% conserved"
7767. .8765
repeat_region
/feature="L1M1 repeat: matches -1390. -.399 of consensus"
8750. .9063
repeat_region
/feature="L1M1 repeat: matches 2. .311 of consensus"
9220. .10853
repeat_region
/feature="L1M1 repeat: matches 311. .2350 of consensus"
15437. .15809
repeat_region
/feature="L1M1A7 repeat: matches 5769. .6127 of consensus"
15827. .16052
repeat_region
/feature="L1M1A8 repeat: matches 6064. .6290 of consensus"
16050. .17506
repeat_region
/feature="L1M1A8 repeat: matches 4330. .5934 of consensus"
17774. .18042
repeat_region
/feature="L1M1A2 repeat: matches 6038. .6308 of consensus"
19407. .19722
repeat_region
/feature="HSMAR2 repeat: matches 271. .588 of consensus"
19723. .20013
repeat_region
/feature="AluYa8 repeat: matches 1. .290 of consensus"
20014. .20732
repeat_region
/feature="HSMAR2 repeat: matches 588. .1299 of consensus"
24209. .24278
repeat_region
/feature="35 copies 2 mer ac 87% conserved"
complement(24831. .25022)
misc_feature
/feature="HF1"
/feature="match: STS: Em:G26756"
complement(join(24831. .25222,26334. .26516,28705. .28881,
30282. .30458,31541. .31714,34673. .34858,35327. .35509,
45393. .45572,45681. .45863,47036. .47212,56564. .56740,
58416. .58598,82094. .82270,82719. .82913,87096. .87269,
92540. .92710,94666. .94857,96268. .96344,98371. .98476,
99170. .99355,120158. .120301))
/feature="HF1"
/feature="H factor 1 (complement): match: cDNAs: Em:M12660
Em:X98697 Em:M29009 Em:M17517 Em:X07523 Em:X04697
Em:X68679 Em:M65293 Em:M65292 Em:X98337 Em:M29008
Em:X56210 Em:X56209 Em:M12383 Em:X64877 Em:M29007"
/product="dJ177P10.1.1 (H factor 1 (complement) isoform
1)"
/evidence=not_experimental
complement(24831. .120301)
/feature="HF1"
CDS
complement(join(25020. .25222,26334. .26516,28705. .28881,
30282. .30458,31541. .31714,34673. .34858,35327. .35509,
45393. .45572,45681. .45863,47036. .47212,56564. .56740,
58416. .58598,82094. .82270,82719. .82913,87096. .87269,
92540. .92710,94666. .94857,96268. .96344,98371. .98476,
99170. .99355,120158. .120215))
/feature="HF1"
/feature="match: proteins: Sw:P06909 Sw:P08603 Tr:Q61408
Tr:Q28085 Tr:Q14570 Tr:Q61407 Tr:Q92496 Sw:Q02985
Tr:Q99254 Tr:Q91275"
/codon_start=1
/evidence=not_experimental
/product="dJ177P10.1.1 (H factor 1 (complement) isoform
1)"
/protein_id="CAB70597.1"
/db_xref="GI:6782358"
/translation="MRLAKIICLMLWAICVAEDCNELPPRRNTEILTGSWDQTYPE

```

```

GTQATYKCRPGYRSLGNVIMVCRKGEWVALNPLRKCQRKPCGHPGDTFFGTFTLTGGN
VFYGVKAVYTCNEGQQLLGEINTYRECDDGTNDIPICEVVKCLPVTAPNGKIVSS
AMEPDRYHFGAORFVCSNGYKTEGDEMHCSDDGFWSKPKCEI3SCSPDIVNG
SPIQKLIYKNERFOYKCMNGYVESRGDAVCTESGWRPLPSCCEKSCDNPYIPNGD
YSLRLKHTGDEITYOCNGFYPATRCNTAKCTSTGWIAPARCTLKPCDYPDIKHGG
LIHENMRPYPPVAVGKYYSYCDHEFTPSGSHYDH1HICTQDGHSPAYPLCKICYFP
YLENGYNQHGKRFVQGSIDVACHPGYALPAOTVTCMENGWSPTRCIHVKTCRSK
SSIDIENGFISQYTYALKKAYCKGLGYVADGETSGTITCGKDGMSAOTPCIKS
CDIPIVFNARTKNDTWFKLNDDTDYEDHGYESNTGTGTCGCGYNGWSOLPICYE
RECELPKIDVHLVPDRKDKYKGEVLKFSCKPGFTIVGPN5VOCYHFCGLSDLPDICK
FOVSCGPPPELLNAGNKEKTEKVEYGHSEVVEYCNPRFLMKGPNKIOCVDGEWITLP
VCIYEESTCGDIPLEHCHGAQLSSPPYYGDSVERFCESESTMHGHSITCIHWYTO
LPQCVAQIQLCPPOPPQIPNSHNMTTTLNRYDGERKVSVLCCOENYLIQGEELITCKDGR
WOSIPLVCVEKIPCSOPPOIEHGTINSSRSQESYAHGTKLSYTCGGFRISENETTC
YMGKWSPPQCEGLPCKSPFISHGVVAHMSDQYQVEEYTKCFEGIDGPAIAKC
LGEKWSHPPICIKTDCLSLSPFENAIAPMGKKDQYKAGEOVTYKCATYYKMDGASNV
CINSRWTCRPTCRDTCNVPTVONAYIVSRQMSKYPSGERVRYOCRSPYEMFGDEEV
MCLANGWTEPPOCKDSTCKCGPPPIDNGDITSPFLSYVAPASSVEYQCONLYQLEGN
KRITCRNGQWSEPPKCLHPCVISREIMENYIALRWTAQKLSYTSVEGESVFCARGY
RLSKRSHTLRTTCWDGKLEYPYTCAR"
25148. .25518
/feature="match: GSS: Em:AQ572291"
26837. .26998
repeat_region
/feature="L1M4c repeat: matches 1281. .1443 of consensus"
27049. .28155
repeat_region
/feature="L1M4c repeat: matches 1537. .2293 of consensus"
complement(28990. .29574)
/feature="HF1"
/feature="match: GSS: Em:AQ322856"
29070. .29574
misc_feature
/feature="match: GSS: Em:AQ309855"
29070. .29462
misc_feature
/feature="match: GSS: Em:AQ015555"
32503. .34070
repeat_region
/feature="THEIC-internal repeat: matches 1. .1580 of
consensus"
35512. .35633
repeat_region
/feature="61 copies 2 mer aa 61% conserved"
36500. .36733
repeat_region
/feature="117 copies 2 mer tt 54% conserved"
37623. .37757
repeat_region
/feature="L1M3d repeat: matches 252. .385 of consensus"
37792. .37944
repeat_region
/feature="L1M3d repeat: matches 508. .654 of consensus"
37955. .38048
repeat_region
/feature="L1M3b repeat: matches -133. -.39 of consensus"
39598. .42996
repeat_region
/feature="L1M1 repeat: matches 857. .3909 of consensus"
43390. .43463
repeat_region
/feature="37 copies 2 mer ta 67% conserved"
43569. .43618
repeat_region
/feature="25 copies 2 mer gt 94% conserved"
complement(43800. .44290)
misc_feature
/feature="HF1"
/feature="match: GSS: Em:AQ595617"
44323. .44795
misc_feature
/feature="match: GSS: Em:AQ785133"
complement(45113. .45337)
/feature="HF1"
/feature="match: GSS: Em:AQ630668"
46395. .46801
repeat_region
/feature="MSTB repeat: matches 4. .404 of consensus"
47551. .48221
repeat_region
/feature="L1PbB repeat: matches -695. .29 of consensus"
48285. .48548
repeat_region
/feature="L1PbB repeat: matches 82. .346 of consensus"
48556. .48620
repeat_region
/feature="L1PbA repeat: matches -233. -.162 of consensus"
48674. .48853
repeat_region
/feature="L1PbA5-16 repeat: matches 663. .848 of consensus"
48988. .50194
repeat_region
/feature="L1M2 repeat: matches 764. .2360 of consensus"
50222. .50330

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 17:19:59 ; Search time 3911.83 Seconds
(without alignments)
2987.953 Million cell updates/sec

Title: US-09-316-163-2

Perfect score: 866

Sequence: 1 tcgagtcacatgctccacaga.....ttgaattacatttgaaaaa 866

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_gss.*
13: em_gss_hum.*
14: em_gss_inv.*
15: em_gss_pln.*
16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	496.6	57.3	652	9	BB609000
2	491.4	56.7	891	10	BF237071
3	448.4	51.8	842	9	AU122695
4	446.8	51.6	764	10	BI460026
5	446.8	51.6	906	9	AL540473
6	445.6	51.5	903	9	AL531827
7	442.8	51.1	890	9	AU116828
8	430.6	49.7	642	9	BB661130
9	401.2	46.3	722	10	BI760742
10	397.4	45.9	686	10	BG927932
11	378	43.6	988	10	BG288341
12	372.2	43.0	659	10	BG898846
13	370.8	42.8	650	9	A1207474
14	368.6	42.6	643	10	BG897905
15	367	42.4	662	10	BG900945
16	365.8	42.2	614	9	A1613807
17	355.8	41.1	640	10	BG901032

18	343	39.6	606	10	BG898574
19	341.4	39.4	578	9	AA500366
20	326	37.6	567	9	AW951533
21	296.2	34.2	395	9	BB845989
22	257.2	29.7	474	10	BF461587
23	252.8	29.2	760	10	BF027852
24	225.2	26.0	646	10	BG925340
c 25	209	24.1	525	9	AI683097
26	205	23.7	432	10	BG899280
27	201.4	23.3	411	10	BG898857
28	192.6	22.2	408	10	BG897147
29	186.6	21.5	743	9	AU117664
30	181	20.9	663	10	BG924822
31	179.8	20.8	398	10	BG898301
32	179.2	20.7	265	9	AA384423
33	176.6	20.4	377	10	BG897473
c 34	173.2	20.0	595	10	BF395144
c 35	173.2	20.0	626	9	AI009773
c 36	173.2	20.0	631	9	AI170314
37	172.4	19.9	649	10	BG925101
c 38	170	19.6	578	9	AI410851
c 39	168.2	19.4	484	9	AI009489
40	163	18.8	288	9	BB601574
41	148.8	17.2	423	10	BG926928
c 42	146	16.9	423	9	AA859019
c 43	142.8	16.5	364	9	AA945005
44	141	16.3	244	9	BB598679
45	139.6	16.1	300	9	AU098607

ALIGNMENTS

RESULT 1

BB609000

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BB609000 RIKEN full-length enriched, 2 days pregnant adult female
ovary Mus musculus CDNA clone E33009L21 5', mRNA sequence.

BB609000.2 GI:16451137

EST.

house mouse.

Mus musculus

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 652)

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hangaki,T., Hara,A.,

Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,K., Konno,H., Kouda,

M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,

Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,

D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,

Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,

Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

On Dec 6, 2000 this sequence version replaced gi:11564176.

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Muramatsu,M., Sugahara,Y., Shibata,K., Itoh

M., Konno,H., Okazaki,Y., Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagii,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura

S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and


```

QY 61 ttatatattgagcgtgttggtagcagaagattgtaaaaggctctctccaaagagaatt 120
Db 94 TTATATATGGAGCT-TTGTGCGAGCAGAGATTGTAAAGCTCCTCTCCAAAGAGAAAT 152

QY 121 cagaataatcctcaggtctcgtggtctgaacaactatattcagaagcactcaggcaacct 180
Db 153 CAGAAATCTCTCAGGCTCGTGGTCAGACAACTATATCCAGNAGCACCAGGCTACCT 212

QY 181 acaaatgcgcgcctggataccgaacacttgggtactattgtataaagtatgcaagaatggag 240
Db 213 ACAAAATGCCGCCCTGGATACCGAACACTTGGCACATATTGTAAAGATGATCAACAATGGAA 272

QY 241 aatgggtaccttctaaccatcaagatattgctggaaaggccatgtggcctcccgag 300
Db 273 AATGGGTGGGCTCTAACCCATCAGGATATGTGGGAAAAGCCTTGTGGCATCCCGGAG 332

QY 301 acacaccccttgggtcctttagctggcagttggatctgaatttgaatttgggtgcaagg 360
Db 333 ACACACCCCTTTGGGTCTTTAGGCTGGCAGTTGGATCTCAATTTGAGTTTGGTGCAAGG 392

QY 361 ttgtttatcatatgtagaagggtaccactatttaggtgaaattgattaccctggaatgtg 420
Db 393 TTGTTTATACCTGTGATGGGTATCAACATATTAGTGCAAAATTGATTACCGTGAATGTG 452

QY 421 atgcagatgggtggaaccatatttccaatatgtgaattgtgaagtcttgcagtgaa 480
Db 453 GTGCAGATGGGTGGATCAATGATATTCCCACTATGTGAAGTTGTGAAGTGTCTACCTGTGA 512

QY 481 cagaactgagaatggaagaattgtgagtggtgagccgaaccagaccaggaattatt 540
Db 513 CAGAATCGAGAATGGAGAATCGTGAGTGGTGCAG-AGAAACAGACCAGGAACTACTATT 571

QY 541 ttggacaggtgtagcgttcttgaatgcaactccggtcctcaagattggaag 589
Db 572 CAGGACAGTGTGGCGTTTGAATGCAATTTCAGGCTTCAAGATTGAAG 620

RESULT 3
LOCUS AUI22695 MAMMAL Homo sapiens cDNA clone MAMMAL1002920 5', mRNA EST 23-OCT-2000
DEFINITION AUI22695 MAMMAL Homo sapiens cDNA clone MAMMAL1002920 5', mRNA
sequence.
ACCESSION AUI22695
VERSION AUI22695.1 GI:10947411
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 842)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1. .842
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MAMMAL1002920"
/clone_lib="MAMMAL"

FEATURES
source
1. .842

```

```

/tissue_type="mammary gland"
/notes="Vector: pME18SFL3"
BASE COUNT 278 a 124 c 205 g 232 t 3 others
ORIGIN

Query Match 51.8%; Score 448.4; DB 9; Length 842;
Best Local Similarity 78.6%; Pred. No. 2.3e-111;
Matches 536; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 5 gtcaactgctcccgatagatccaagaatgagactgtcagcaagaattatttggccttat 64
Db 46 GTPAATGTCTCTTAAAGATCCAAAATAGACACTTCTAGCAAAAGATTATTTCCTTAT 105

QY 65 attatggaactgttgtgtagcagaagattgtaaaaggctccctcccaagagaaaattcaga 124
Db 106 GTTATGGGCTATTGTGTAGCAGCAAGATGCAATGAACCTTCCTCCAGAGAAATACAGA 165

QY 125 aattctctcaggtctgtgctgaaacaactatattcagaaggcactcaggaacacctacaa 184
Db 166 AATTCCTGACAGGTTCTGTGCTGACCAACATATCCAGAAAGGCCACCCAGGCTATCTATAA 225

QY 185 atgcccgccttgatcacgaacacttgggtactattgtaaaagtatgcaagaatgagaatg 244
Db 226 ATGCCCCCTGGATATAGATCTCTTGGAAATATAATGTTGATGCAGGAAGGAGAATG 285

QY 245 ggtaccttctaaacctcaaggatgtcggaaaagcccatgtggcctcccgagacac 304
Db 286 GGTGTCTCTTAATCCATTAAGGAATGTCAGAAAGGCCCTGTGGACATCTCTGGAGATAC 345

QY 305 accttgggtgctttaggtggtgagctggtgagctggaatttgggtgcaagggtgtg 364
Db 346 TCGTTTGGTACTTTTACCCCTTACAGGAGGAATGTGTTTGAATATGCTGTAAAGCTGT 405

QY 365 ttatacatgtgatgaagggtacacactattaggtgaaattgattaccggaagtgtgac 424
Db 406 GTATACATGTAATGAGGGGTATCAATGCTAGGTGAGATTAATACCGTGAATGTGCAC 465

QY 425 agatgggtggacaaatgatattccaatatgtgaagtgtgaaagtgtgcagtgacaga 484
Db 466 AGATGGATGGACCAATGATATTCCTATATGTGAAGTTGTGAAGCTTTTACCAGTGACAGC 525

QY 485 actggagaatggaagaattgtgagtggtgagccgcaaccagaccaggaatattatttgg 544
Db 526 ACCAGAGAATGGAAAAATTTGTCAAGTAGTGCATGCAATGGAACCCAGATCGGGAATTTGG 585

QY 545 acaggtggtacgttctgaaatgcaactccgcttcaagattgaaagacagaaagaatgca 604
Db 586 ACAAGCAGTACGGTTTGTATGTAACTCAGGCTACAAAGTTGAAGGAGATGAAGAAATGCA 645

QY 605 ctgctcataaaatggcctctggagcaatgaaaagccacagtggtggaattttcttgcct 664
Db 646 TTGTTTCAGACGATGGTTTGGAGTAAACAGAAACCAACCAAGTGTGTGGAATTTTCATGCAA 705

QY 665 gccacacaggtgaaatcaga 686
Db 706 ATCCCCAGATGTTATAAATGGA 727

RESULT 4
LOCUS BI460026
DEFINITION BI460026 603201565F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5267249 5',
mRNA sequence.
ACCESSION BI460026
VERSION BI460026.1 GI:15250682
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 764)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

```



```

|||||
Db 210 ATGCCCCCTGGATATAGATCTCTTGGAAATGTAATTAATGGTATGAGGAAGGAGAATG 269
QY 245 ggtacaccttaaccatcaagcatgtcggaagccatgtggcaccatcccgagacac 304
Db 270 GGTGTCTTAAATCCATTAAGGNAATGTGAGAAAGGCCCTGTGGACATCCTGGAGATAC 329
QY 305 accctttgggtcccttaggctggcagtgatcgatgaattgaatttggtgcaaggttgt 364
Db 330 TCCTTTTGGTACTTTTACCCCTTACAGGAGGAAATGTCTTTGAATATGGTCTAAAGCTGT 389
QY 365 ttatcatatgatgaaggtaccacattattagtgaaattgattaccgtgaattgtatgc 424
Db 390 GTRATACATGTAATGACGGGTATCAATGCTAGGTGAGATTAATTACCGTGAATGTGACAC 449
QY 425 agatgggtggaccaatgatattccaatatgtgaagttggaagtgcctgpcagtgacaga 484
Db 450 AGATGATGACCAATGATATTCCTATATGTGAAGTTGTGAAGTGTACCAGTGACAGC 509
QY 485 actggagaatgaagaattgtgagtggtgcagccgaaccagaccaggaatatattttgg 544
Db 510 ACCAGAGAATGGAATAATTTGTAGTGTGCAATGGAACCCAGATCGGGAATACCATTTTGG 569
QY 545 acaggtggtacgtttggaatccagctccgcttcaagattgaaggaagacagaaagatgca 604
Db 570 ACNAGCAGTACCGTTGTATGTAATCACTAGGCTACAGATTTGAAGGAGATCAAGAAATGCA 629
QY 605 ctgctcataaaatggcctctggagcaatgaaagccacagtggtgtggaattcttgcct 664
Db 630 TTGTTTCAGACGATGGTGTGTTGGAGTAAAGAGAAACCAAAAGTGTGTGGAATTTTCATGCAA 689
QY 665 gccaccacagagttgaaatgga 686
Db 690 ATCCCCCAGATGTTATAAATGGA 711

```

```

RESULT 6
AL531827 903 bp mRNA linear EST 13-FEB-2001
LOCUS AL531827 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DM003YM12 5
DEFINITION prime, mRNA sequence.
ACCESSION AL531827
VERSION AL531827.1 GI:12795320
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..903
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DM003YM12"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610

```

```

8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com*
BASE COUNT 299 a 139 c 204 g 253 t 8 others
ORIGIN
Query Match 51.5%; Score 445.6; DB 9; Length 903;
Best Local Similarity 78.2%; Pred. No. 1.3e-110;
Matches 533; Conservative 2; Mismatches 147; Indels 0; Gaps 0;
QY 5 gtcaactgctcccgatagatccaagaatgagactgacgaagaattatttggcttat 64
Db 19 GTAATGTCTCTTAAAGATCCAAAATGAGACTTCTAGCAAAAGATTATTTCCTTAT 78
QY 65 attatggaactgtttgttagcagaagattgtaaggctccctccccaagagaaattcga 124
Db 79 GTTATGGCTATTGTGTACGACAAGATTCGAATGAACCTTCTCCCAAGAGAAATACA 138
QY 125 aattctctcaaggttcgtggtctgaaacaactatatccaagaggcaactcaggcaacctacaa 184
Db 139 AATTCTGACAGGTTCTCGTGTCTGACCAACATATCCAGAAGGCCACCCAGGCTATCTATAA 198
QY 185 atgcgccttggaataccgaacacttggtactattgtataaagtatgcaagaatggagaatg 244
Db 199 ATGCCGCCCTGGATATAGATCTCTTGGAAATGTAATAATGTTATGCDGGRAGGAGAATG 258
QY 245 ggtacaccttaaccatcaaggtatgtcggaagccatgtggcaccatcccgagacac 304
Db 259 GGTGTCTTAAATCCATTAAGGAAATGTGAGAAAGGCCCTGTGGACATCCTGGAGATAC 318
QY 305 accctttgggtcccttaggctggcagtgatcgatgaatttggtgcaaggttgt 364
Db 319 TCCTTTTGGTACTTTTACCCCTTACAGGAGGAAATGTCTTTGAATATGTTGTAAGCTGT 378
QY 365 ttatcatatgatgaaggtaccacacttattagtgaaattgattaccgtgaattgtatgc 424
Db 379 GTRATACATGTAATGAGGGTATCAATGCTAGGTGAGATTAATTACCGTGAATGTGACAC 438
QY 425 agatgggtggaccaatgatattccaatatgtgaagttgtgaaagtgccttgcagtgacaga 484
Db 439 AGATGATGACCAATGATATTCCTATATGTGAAGTTGTGAAGTGTACCAGTGACAGC 498
QY 485 actggagaatgaagaattgtgagtggtgcagccgaaccagaccaggaatatattttgg 544
Db 499 ACCAGAGAATGGAATAATTTGTAGTGTAACTCAGGCTACAAGATTGAAGGAGATGAAGAAATGCA 558
QY 545 acaggtggtacgtttggaatccagctccgcttcaagattgaaggaagacagaaagatgca 604
Db 559 ACNAGCAGTACCGTTGTATGTAATCACTCAGGCTACAAGATTGAAGGAGATGAAGAAATGCA 618
QY 605 ctgctcataaaatggcctctggagcaatgaaagccacagtggtgtggaattttctgcct 664
Db 619 TTGTTTCAGACGATGGTGTGTTGGAGTAAAGAGAAACCAAAAGTGTGTGGAATTTTCATGCAA 678
QY 665 gccaccacagagttgaaatgga 686
Db 679 ATCCCCCAGATGTTATAAATGGA 700

```

RESULT 7

```

AL116828 890 bp mRNA linear EST 19-OCT-2000
LOCUS AL116828 HEMBA1 Homo sapiens cDNA clone HEMBA1000088 5', mRNA
DEFINITION sequence.
ACCESSION AL116828
VERSION AL116828.1 GI:10931750
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 890)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,

```

Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and Isogai, T.
 HRI human cDNA project
 Unpublished (2000)
 Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp
 HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.
 Location/Qualifiers
 1. .890
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEMBA1000088"
 /clone_lib="HEMBA1"
 /tissue_type="whole embryo, mainly head"
 /dev_stage="embryo, 10 weeks"
 /note="Vector: pME18SFL3"
 BASE COUNT 298 a 136 c 213 g 239 t 4 others
 ORIGIN

Query Match 51.1%; Score 442.8; DB 9; Length 890;
 Best Local Similarity 78.3%; Pred. NO. 7.7e-110;
 Matches 531; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 5 gtaactgtctccagatagatccaagacatgagactgtcagcaagaattatttggcttat 64
 Db 49 GTAATGTCCTCTTAAAGATCCAAAATGACACTCTAGCAAAAGATTATTGCCCTTAT 108
 QY 65 attatgactgttctgttagcagaagaattgtaaagtctctctcccaagagaaattcaga 124
 Db 109 GTTATGGGTATTGTGTAGCAGAAGATTGCAATGAACCTCTCCCAAGAGAATACAGA 168
 QY 125 aattctctcaggcttgctgtaacaactattatcagaagcaactcaggcaacctacaa 184
 Db 169 AATTCTGACAGGTTCTGTGTCGACCAACATATCCAGAAAGGCCCGCTATCTATATA 228
 QY 185 atgccccttggtataccggaacacttggtactattgttaaagtatgcaagaatggagaattg 244
 Db 229 ATGCCCGCTGGATATAGATCTCTTGGAAATATGATAATGGTATGTCAGGAAGGGAATC 288
 QY 245 ggtaccttcaaccatacgaagatatgtcggaaaaggccatgtgggcatcccgagacac 304
 Db 289 GGTGTCTCTTAATCCATTAAAGGAATGTGAGAAAAGGCCCTGTGGACATCCTGGAGATAC 348
 QY 305 acccttgggtccctttaggctggcagttggtatctggaattggaattgggtgcaaaagtgtt 364
 Db 349 TCCTTTTGGTACTTTTACCCTTACAGGAGGAATGTGTTGAATATGTTGTTAAACCTGT 408
 QY 365 ttatacatgcatgaagggtacaaactatagtgtaaatgattaccgtgaattgtgatgc 424
 Db 409 GTATACATGTAATGAGGGGTATCAATTTGTAGGTGAGATTAATTACCGTGAATGTGACAC 468
 QY 425 agatgggtgaccaatgatattccaatatgtgaagtgtgaagtgttccagtgacaga 484
 Db 469 ACATGATGGACCAATGATATTCCTATATGTGTAGTGTGAAAGTGTTTACCAAGTGACGC 528
 QY 485 actggagaatggaagaattgtgagtgtgagccgacccagaccaggaataattatttgg 544
 Db 529 ACCAGAGATGGAATAATTTTCAGTAGTCAATGGAAACAGCATCGGAATACCATTTTCG 588
 QY 545 acaggtgtgaactcttgaatgaactccggcttcaagattggaagacagagaagaatgca 604
 Db 589 ACAAGCAGTACGGTTTGTGTGTAATCTACCTACGCTACAAGATTGAAGGATGAAGAAATGCA 648
 QY 605 ctgtcctataaatggcctctggagcaatgataaagccacagtggtggaatttcttgcct 664

Db 649 TTGTTCCAGACGATGGTTTTTGGAGTAAGAGAAACCAAGTGTGTGGAATTTTCATCAA 708
 QY 665 gccaccacagagttgaaaa 682
 Db 709 ATCCCCAGATGTTATATA 726

RESULT 8
 BB661130
 LOCUS
 DEFINITION
 musculus cDNA clone D630023013 5', mRNA sequence.
 ACCESSION
 BB661130
 VERSION
 BB661130.1
 KEYWORDS
 EST.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 642)
 REFERENCE
 AUTHORS
 Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,
 Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda,
 M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,
 Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki,
 D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
 Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyota, T.,
 Muramatsu, M. and Hayashizaki, Y.
 Riken Mouse ESTs (Arakawa, T., et al. 2001)
 Unpublished (2001)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayashi, N., Sugahara, Y., Shibata, K., Itoh,
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 Wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura,
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara,
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yananaka, I., Aizawa,
 K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
 Hayashizaki, Y.
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.
 e mouse tissues.
 Location/Qualifiers
 1. .642
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="D630023013"
 /clone_lib="RIKEN full-length enriched, 0 day neonate
 kidney"
 /tissue_type="kidney"
 /dev_stage="0 day neonate"

FEATURES
 source

1 (bases 1 to 722)

```

/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCAACTCGAGTTTTTTTTTTTTVN 3']. cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCTCGATTAAATTAATTCCTCCCCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+ ) after bulk excision from Lambda FLC I."
177 a 127 c 156 g 182 t
BASE COUNT

```

[illegible][illegible]

```

Qy 489 gagaatggaagaattgtgagtggtgcagccgaacacagaccagggaatattattttggacag 548
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 521 GAGATGGAAGAAATGTGCTAGTAGTGCATGGAACACAGATCGGAATACCATTTTGGACAA 580
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 549 gtggtacgctttgaatcaactcgcgttcagattgaagacagagaagaataatgcactgc 608
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 581 CGATGCGGTTGTATGTAACCTACAGCTACAGGATACAGGATGACGGAGATGAAGAAATGCATGT 640
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 609 tcataaatgagcct-citggagcaatgaaagccagctgtgtggaattt 657
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 641 TCACAGCATGGTTGATGGAGTAGAAGAGAAACCAAGTGTGTGGAACATT 690
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
BG927932 686 bp mRNA linear EST 06-NOV-2001
LOCUS HNC66-1-H3.R HNC (Human Normal Cartilage) Homo sapiens cDNA, mRNA
DEFINITION
sequence.
ACCESSION BG927932
VERSION BG927932.1 GI:14322455
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 686)
Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
Sathe,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
Lark,M.W.
IDENTIFICATION AND INITIAL CHARACTERIZATION OF 5000 EXPRESSED
SEQUENCED TAGS (ESTs) EACH FROM ADULT HUMAN NORMAL AND
OSTEOARTHRIC CARTILAGE CDNA LIBRARIES
Osteoarthritis Cartilage 9 (7), 641-653 (2001)
21482651
Contact: Sanjay Kumar
UW2109
GlaxoSmithKline
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
Tel: 610-270-7245
Fax: 610-270-5598
Email: sanjay_kumar-l@sk.com
Seq primer: T7.
Location/Qualifiers
1..686
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HNC (Human Normal Cartilage)"
/tissue_type="cartilage"
/lab_host="E.coli DH10 B"
/notes="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
Directional"
BASE COUNT 214 a 126 c 167 g 179 t
ORIGIN
Query Match 45.9%; Score 397.4; DB 10; Length 686;
Best Local Similarity 79.5%; Pred. No. 1.6e-97;
Matches 470; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

Qy 2 caggtcaactgcctccagatagatccaagacatgagactgcagcagaagaattattggct 61
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94 CGGGTAAATGTCCTCTTAAAGATCCAAAATGAGACTTCTAGCAAGATTATTTCCT 153
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 62 tatattatgactgtttgttagcagaagattgttaaggtcctccccaagaaaattc 121
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 154 TATGTTATGGCTTATTGTGTACAGAAGATTGCAATGAACTTCTCTCCAAAGAAATAC 213
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 122 agaaattctctcaggtcgtggtcgaacactatattcagaagggcactcagcaacct 181
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 214 AGAATCTCACAGGTTCTGGTCTGACCAACATATCCAGAAGGCCACCCAGGCTATCTA 273
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 182 caaatgccccctggatccgaacacttggtactattgttaaaagtatgaagaatggaga 241
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 274 TAAATGCCGCCCTGGATATAGATCTCTTTGGAAATATAATAATGGTATGCAGGAAGGAGA 333
Qy 242 atgggtaccttctaccatcaaggaatgtcggaaaagccatgtggcaccctccggaga 301
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 334 ATGGGTGCTCTTAATCAAGAAATGTCAGAAAAGGCCCTGTGGACATCTCTGGAGA 393
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 302 cacaccctttgggtcctttagctggcagttggactcctgaatttgaatttgggtgcaaaagt 361
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 394 TACTCCTTTTGGTACTTTTACCCTTACAGGAGGAATATGTTTGAATATGGTGTAAAGC 453
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 362 tgtttatcatgtgatgaaggggtaccactattaggtgaaattgattccctggaatgga 421
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 454 TGTGTATACATGAATGAGGGGTATCAATGCTAGGTAGATTAATATCCGTGAATGTGA 513
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 422 tgcagatgggtggacccaatgatattccaatatgtgaagtgtgaagtgtgcagtcgtccagtgac 481
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 514 CACAGATGGATGGACCAATGATATCTTATATGTGAAGTGTGGAAGTGTTTACCACTGAC 573
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 482 agaactggagaatggagaatgtgagtggtgcagccgaacacagaccagggaatattatt 541
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 574 AGCACCAGAGAATGGAAAATTTGTCAGTAGTGCATGCAATGCAACAGATCGGAATACCATTT 633
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 542 tggacaggtgtacgttttgaatgcaactcgcgttcagatttgaagattgaaggaga 592
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 634 TGGACAAGCAGTACGGTTTGTGTGTAACCTCAGGCTTACAAGATTGAAGGAGA 684
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
BG288341 988 bp mRNA linear EST 21-FEB-2001
LOCUS 602383678F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4512906 5',
DEFINITION mRNA sequence.
ACCESSION BG288341
VERSION BG288341.1 GI:13043080
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 988)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
UNPUBLISHED (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10398 row: k column: 19
High quality sequence stop: 661.
Location/Qualifiers
1..988
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4512906"
/clone_lib="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/notes="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 336 a 134 c 259 g 258 t
ORIGIN
Query Match 43.6%; Score 378; DB 10; Length 988;
Best Local Similarity 76.2%; Pred. No. 3.6e-92;

```

Matches 491; Conservative 0; Mismatches 150; Indels 3; Gaps 2;

```

QY 5 gtaactgtctccagatagatccaagacatgagactgtcagcaagaattatttggcttat 64
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 28 GTAATGTCTCTTAAAGATCCAAAAATGAGACTTCTAGCAAAAGATTATTGGCTTAT 87
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 65 attagagactgttgttagcagaagaattgtaaggctcctcccaagagaaaaattcaga 124
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 88 GTTATGGGCTA-TTGTGTACGAGAAGATTGCAATGAAGACTTCTCCCAAGAGAATAACAGA 146
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 125 aattctctcaggctcgtggtctgaaacaactatattcagaaggcactcaggcaacctacaa 184
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 147 AATTCTGACAGGTTCTGTGCTGACCAACATATCCAGAGGACCCAGGCTATCTATAA 206
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 185 atgccccttggtatccgaacaacttggtaactatgttaaaagtgtcaagaatggagaatg 244
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 207 ATCCGCCCTGGATATAGATCTCTTGGAAATGTAATAATGCTATGCAAGGAGGAGAATG 266
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 245 ggtacctcttaacccatcaggatattgcgaaagccatgtggcgcaccccgagacac 304
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 267 GGTGTCTTTAATCCATTAGGAAATGTCAGAAAGGCCCTGTGGACATCCTGGAGATAC 326
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 305 acccttgggtccttagtgcggtcagttggtatgtaatttgaatttgggtcaagggtgtg 364
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 327 TCCTTTTGTGTACTTTACCTTACAGGAGGAATGCTTTTGAATATGCTGTAAAGCTGT 386
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 365 ttatacatgtgatgaagggttaccacactattagggtgaattgattaccgtgaattgtatgc 424
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 387 GTATACATGTAATGAGGGGTATCAATTTGCTTAGGTGAGATTAAATCCGCTGAATGTGAC 446
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 425 agatgggtgagcaaatgatattccaatatattgaatttgaatttgggtcagtgacaga 484
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 447 AGATGATGACCAACATGATATCTTATGTGAAGTTGTGAAGTGTGTACACGTGACAGC 506
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 485 actggagaatggaagaattgtgagtggtgcagccgaacacagaccaggaattatttttgg 544
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 507 ACCAGAGAATGGAAAAATGTCAGTAGTGAATGGAACACAGATCGGGAATACCATTTGGA 566
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 545 acagggtggtacgttttgaattgcaactccggcttcagattggaaggacagaagaatgca 604
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 567 CAAGCAGTAGTACGG--TTGTATGTAACTCAGGCTAAAGATTGAAGGAGATGAAGAAATGCA 624
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 605 ctgctcataaattgacctcgtggagcaatgaaagccacacagtgtg 648
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 625 TTGTTTACAGCATGGTTTGGAGTAAGAGAAACCAAGTGTGTG 668
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 12

LOCUS BG988846 659 bp mRNA linear EST 06-NOV-2001
 DEFINITION HOA4-1-D-8 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA, mRNA sequence.

ACCESSION BG988846
 VERSION BG988846.1 GI:14309095

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 659)
 AUTHORS Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J., Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Lark,M.W.

TITLE Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritic cartilage cDNA libraries

JOURNAL Osteoarthritis Cartilage 9 (7), 641-653 (2001)

MEDLINE 21482651

COMMENT Contact: Sanjay Kumar

UW2109

GlaxoSmithKline

709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA

Tel: 610-270-7245

Fax: 610-270-5598

Email: sanjay_kumar-1@sk.com

Seq primer: T7.

Location/Qualifiers

source

1..659

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="HOA (Human Osteoarthritic Cartilage)"

/tissue_type="cartilage"

/lab_host="E.coli DH10 B"

/note="Vector: pSPORT 1; Site_1: SalI; Site_2: NotI;

Directional"

BASE COUNT 204 a 116 c 162 g 177 t

ORIGIN

Query Match 43.0%; Score 372.2; DB 10; Length 659;
 Best Local Similarity 79.0%; Pred. No. 1.2e-90;
 Matches 443; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 5 gtaactgtctccagatagatccaagacatgagactgtcagcaagaattatttggcttat 64

Db 98 GTAATGTCTCTTAAAGATCCAAAAATGAGACTTCTAGCAAAAGATTATTGGCTTAT 157

QY 65 attagagactgttgttagcagaagaattgtaagggtcctcctcccaagagaaaaattcaga 124

Db 158 GTTATGGGCTATTGTGTAGCAGAAGATTGCAATGAAGACTTCTCCCAAGAGAATAACAGA 217

QY 125 aattctctcaggctcgtggtctgaaacaactatattcagaaggcactcaggcaacctacaa 184

Db 218 AATTCTGACAGGTTCTGTGCTGACCAACATATCCAGAGGACCCAGGCTATCTATAA 277

QY 185 atgccccttggtatccgaacaacttggtaactattgttaaaagtgtcagaagaatggagaatg 244

Db 278 ATCCGCCCTGGATATAGATCTCTTGGAAATGTAATAATGCTATGCAAGGAGGAGAATG 337

QY 245 ggtacctcttaacccatcaggatattgcgaaagccatgtgggcaccccgagacac 304

Db 338 GGTGTCTTTAATCCATTAGGAAATGTCAGAAAGGCCCTGTGGACATCCTGGAGATAC 397

QY 305 acccttgggtccttagtgcggtcagttggtatgtaatttgaatttgggtcgaagggtgtg 364

Db 398 TCCTTTTGTGTACTTTTACCTTACAGGAGGAAATGTGTTGAATATGTTGTAAAGCTGT 457

QY 365 ttatacatgtgatgaagggtaccacactattagggtgaattgattaccgtgaattgtgatgc 424

Db 458 GTATACATGTAATGAGGGGTATCAATTTGCTTAGGTGAGATTAAATCCGCTGAATGTGACAC 517

QY 425 agatgggtggaaccaatgatattcccaatattgtaagtgtgtaagtgtggtccagtgacaga 484

Db 518 AGATGGATGGACCAATGATATCTCTATATGTGAAGTTGTGAAGTGTGTACCAAGTGACAGC 577

QY 485 actggagaatggaagaattgtgagtggtgcagccgaacacagaccaggaataattatttgg 544

Db 578 ACCAGAGAATGGAAAAATTTGTCAGTAGTGAATGGAAGGAGATCGGGAATACCAATTGG 637

QY 545 acagggtgtacgcttttgaattg 565

Db 638 ACAAGCAGTACCCGTTGTATG 658

RESULT 13

LOCUS AI207474 650 bp mRNA linear EST 11-NOV-1999

DEFINITION HA2827 Human fetal liver cDNA library Homo sapiens cDNA, mRNA sequence.

ACCESSION AI207474

VERSION AI207474.1 GI:6361482

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE 1 (bases 1 to 650)
AUTHORS Yu,Y., Zhang,C., Luo,L., Ouyang,S., Li,W., Wu,J., Zhou,S., Liu,M.
          and He,F.
TITLE Expression profile analysis of a human fetal liver cDNA library
JOURNAL Unpublished (1998)
COMMENT Contact: Chenggang Zhang
          Beijing Institute of Radiation Medicine
          27 Taiping Road, Beijing 100850, P.R.China
          Email: zhang_chenggang@hotmail.com.

FEATURES
source Location/Qualifiers
          1..650
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone_lib="Human fetal liver cDNA library"
          /tissue_type="liver"
          /dev_stage="fetal"
          /lab_host="MC1061/p3"
          /note="Vector: pCDNA1"

BASE COUNT 208 a 102 c 158 g 182 t
ORIGIN

Query Match 42.8%; Score 370.8; DB 9; Length 650;
Best Local Similarity 76.7%; Pred. No. 2.8e-90;
Matches 480; Conservative 0; Mismatches 142; Indels 4; Gaps 2;

Qy 5 gtcaactgtctccagatagatcccaagacatgagactgtcagcaagaattatttggcttat 64
Db 18 GTAATGTCCTTTAAAGATCCAAAATGAGACTTCTAGCAAGAAATTATTTCCTTAT 77

Qy 65 attatgactgtttgttagcagaagattgaaaggctctctcccaagagaaattcaga 124
Db 78 GTTATGGCTATTTGTGTAGCAGAAGATTGCAATGAATCTCTCCCAAGAGAAATACAGA 137

Qy 125 aattctctcagggttcgtgtcgaacactatttcagaagccactcaggcaacctcaaa 184
Db 138 AATTCTCACAGGTTCCTGGTCTGACCAACATATCCAGAAAGCCACCCAGGCTATCTATA 197

Qy 185 atgcgcctctggataccgaacacttggtactattgttgaagatgcgaagaatggagaatg 244
Db 198 ATGCGCGCTGGATATAGATCTCTGTGAAATGTAATAATGATGTCAGGAAGGAGAATG 257

Qy 245 ggtaaccttaaccacaaagatagatgcgaaagccatgtggccatcccgagagacac 304
Db 258 GGTGTCTCTAATCCATTAGGAATGTCAGAAAGGCCCTGTGGACATCTCGAGATAC 317

Qy 305 acccttgggtccctttaggtggcagttggatctgaattggaattgtgtgcgaagggtgt 364
Db 318 TCCTTTTGGTACTTTTACCCTTACAGGAGGAAATGCTTTGAATATGTTGTAAGCTGT 377

Qy 365 ttatactgtgaaaggggtaccaactatttaggtgaaattgattaccgtgaaatgtgatgc 424
Db 378 GTATACATGTAATGAGGGGTATCAATTTGCTAGGTGAGATTAAATTTACCGTGAATGTGAC 437

Qy 425 agatgggtggaccacaa----tgattccaatatgtgaagttgaaagtcgttcgcaagtgc 481
Db 438 AGATGGATGTCACCAATGATATATCTCTATATGTGAAGCTGCGAAGTGTTCACCGTGC 497

Qy 482 agaactggagaatggaagaattgtgagtgagcgcgcaaccagacagcaggaattatttt 541
Db 498 AGTACCAGAGAAATGGAATAATTTGTCAGTGTGCAATGGAACCCAGATCGGGAATACCAATT 557

Qy 542 tggacaggtggtacg-ctttgaatgaactccggcttcaagattgaaggacagaaagaaa 600
Db 558 TGGACAAAGCAGTACGTTGTTGTATGTAACTCAGGCTACAGGCTACAAAGATTGAAGGAGATGAAGAAA 617

Qy 601 tgcactgctcataaaatggcctctgg 626
Db 618 TGCATTGGCTCAGACCATGGTTTTTTG 643

RESULT 14
BG897905

```

```

LOCUS BG897905
DEFINITION HOA26-1-84 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,
          mRNA sequence.
ACCESSION BG897905
VERSION BG897905.1 GI:14308154
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 643)
AUTHORS Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
          Sathe,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
          Lark,M.W.
TITLE Identification and initial characterization of 5000 expressed
          sequenced tags (ESTs) each from adult human normal and
          osteoarthritic cartilage cDNA libraries
JOURNAL Osteoarthritis Cartilage 9 (7), 641-653 (2001)
MEDLINE 21482651
COMMENT Contact: Sanjay Kumar
          UW2109
          GlaxoSmithKline
          709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
          Tel: 610-270-7245
          Fax: 610-270-5598
          Email: sanjay_kumar-l@sk.com
          Seq primer: T7.
          Location/Qualifiers
          1..643
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone_lib="HOA (Human Osteoarthritic Cartilage)"
          /tissue_type="cartilage"
          /lab_host="E.coli DH10 B"
          /note="vector: pSPORT 1; Site_1: SalI; Site_2: NotI;
          Directional"

BASE COUNT 197 a 119 c 158 g 169 t
ORIGIN

Query Match 42.6%; Score 368.6; DB 10; Length 643;
Best Local Similarity 79.9%; Pred. No. 1.1e-89;
Matches 434; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 20 atagatccaagacatgagactgtcagcaagaattatttggcttattatgactgtttg 79
Db 95 AAAGATCCAAAATGAGACTTCTAGCAAGAAATTATTTCCTTATGCTTATGGCTATTG 154

Qy 80 tgtagcgaagattgtaaaggtcctctcccaagagaaattcagaattctctcaggttc 139
Db 155 TGTAGCAGAGATTGCAATGAATTCCTCCAAAGAAATACAGAAATCTGACAGGTTTC 214

Qy 140 gtggtctgaacaactatttcagaagcactcaggcaacctcaaaaatgccccctggata 199
Db 215 CTGCTCTGACCAACATATCCAGAAAGCCACCCAGGCTATCTATAATGCCCCCTGGATA 274

Qy 200 ccgacacttggtactattgttaaagtatgcaagaatggagaatgggtgacctcttaaccc 259
Db 275 TAGATCTCTTGGAAATGTAATAATGTTATGCAGGAGGGAGAAATGGGTGTCTCTTAATCC 334

Qy 260 atcaaggatattgtcgaaaggccatgtgggcatccccggagacacaccccttgggtcctt 319
Db 335 ATTAAGGAAATGTCAGAAAGGCCCTGTGGACATCTCTGGAGATACTCTCTTTGGTACTTT 394

Qy 320 taggctggcagttggtatctgaattgattggtgcaaaaggtttttatatacatgtatga 379
Db 395 TACCCTTACAGGAGGAAATGTGTGAATATGTTGTTAAAAGCTGTGTATACATATAATGA 454

Qy 380 agggtaaccaactattaggtgaaattgattaccgtgaaatgtgatcagatgggtgaccaa 439
Db 455 GGGGTATCAATTGCTAGTCAGATTAATTACCGTGAATGTGACACAGATGGATGGACCAA 514

Qy 440 tgatatccaatatgtgaagtgtgtgaagtgtgtgccagtgacagaactggagaatggaag 499

```


GenCore version 4.5
 Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 19:15:26 ; Search time 537.93 Seconds
 (without alignments)
 2764.017 Million cell updates/sec

Title: US-09-316-163-2
 Perfect score: 866
 Sequence: 1 tcgagtaactgtctccaga.....ttgaattacatttgaaaaa 866

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : N_Geneseq_032802.*

	1:	2:	3:	4:	5:	6:	7:	8:	9:	10:	11:	12:	13:	14:	15:	16:	17:	18:	19:	20:	21:	22:	23:	24:
	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	432.4	49.9	525	21 AAA43501	Mouse secreted exp
2	218.8	25.3	649	18 AAV02790	Human partial comp
3	217.2	25.1	581	18 AAV02791	Clone PR89PH410 C
4	46.2	5.3	3696	22 AAH75787	Receptor 222 codin
5	44.6	5.2	2539	21 AAC77331	Human ORFX ORF2886
6	44.6	5.2	2329	22 AAS03887	Human secreted pro
7	44.6	5.2	3804	22 ABA06548	Human cDNA SEQ ID
8	44.6	5.2	3804	22 AAS28968	cDNA encoding for
9	44.6	5.2	3804	22 AAS26868	Human cDNA encodin

ALIGNMENTS

RESULT 1

AAA43501 ID AAA43501 standard; cDNA; 525 BP.

XX AAA43501;

XX 21-AUG-2000 (first entry)

DE Mouse secreted expressed sequence tag SEQ ID NO:76.

XX Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;

KW expressed sequence tag; EST; probe; chemotactic; proliferative;

KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;

KW thrombolytic; antiflammatory; cytostatic; antibacterial; antifungal;

KW antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;

KW antitumor; osteoprotective; neuroprotective; nootropic; antiprosiatric;

KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;

KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;

KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;

KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;

KW central nervous system disorder; Alzheimer's disease; stroke;

KW Parkinson's disease; Huntington's disease; coagulation disorder;

KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;

XX tumour; infection; depression; psoriasis; ss.

XX Mus musculus.

XX WO200021991-A1.

XX 20-APR-2000.

XX 15-OCT-1999; 99WO-US24206.

	10	44.6	5.2	3804	22	AA531587	cDNA encoding nove
	11	44.6	5.2	4088	22	AAK94919	Human full-length
	12	44.6	5.2	5421	22	AAK94920	Human full-length
	13	44.6	5.2	6153	24	ABA03880	Human POLY11 nucle
	14	42.6	4.9	6131	24	ABL32890	Human immune syste
	15	42.2	4.9	372	16	AAT23604	Human gene signatu
	16	41.6	4.8	1430	22	AHH33539	Human colon cancer
	17	41.6	4.8	2519	19	AAV54124	Human membrane pro
	18	41.6	4.8	3139	21	AAC75833	Human ORFX ORF1388
	19	41.6	4.8	14781	22	AAI36303	Human musculoskele
	20	40	4.6	1878	22	AAI60621	Human polynucleoti
	21	40	4.6	1927	22	AAI58835	Human polynucleoti
	22	40	4.6	2404	22	AAH15794	Human cDNA sequenc
	23	40	4.6	4590	22	AAH24065	Yeast AOD9604-asso
	24	39.8	4.6	6131	24	ABL32891	Human immune syste
	25	39.4	4.5	49999	20	AA223902	Human LOBO homolog
	26	38.4	4.4	1640	17	AAT12162	Partial pJG4-5-CDK
	27	37	4.3	6390	23	ABL12023	Drosophila melanog
	28	37	4.3	12519	23	ABL12022	Drosophila melanog
	29	36.8	4.2	1151	21	AAC40206	Arabidopsis thalia
	30	36.8	4.2	2006	21	AAC42229	Arabidopsis thalia
	31	36.2	4.2	385	18	AAV02797	Clone pZS3PH2576 #
	32	36.2	4.2	385	18	AAV02798	Clone pZS3PH2576 #
	33	36.2	4.2	771	22	AAK93472	Human cDNA clone r
	34	36.2	4.2	14924	24	ABL32225	Human immune syste
	35	35.6	4.1	1747	20	AAZ42234	Human normal bladd
	36	35.6	4.1	1765	22	AAK94846	Human full-length
	37	35.6	4.1	1880	22	AAH48345	Human granule memb
	38	35.6	4.1	5409	21	AAA70151	Plasmodium falci
	39	35.4	4.1	183	22	ABA74572	Human foetal liver
	40	35.4	4.1	183	22	AAK23041	Human brain expres
	41	35.4	4.1	183	22	AAK49218	Human bone marrow
	42	35.4	4.1	183	22	AAI5053	Probe #23739 used
	43	35.4	4.1	515	22	ABA62078	Human foetal liver
	44	35.4	4.1	515	22	AAK10395	Human brain expres
	45	35.4	4.1	515	22	AAK36298	Human bone marrow

SQ Sequence 525 BP; 150 A; 101 C; 131 G; 142 T; 1 other;

[illegible]

QY 489 ggaatggaagaattgtgagtgccagccgaaccagaccaggaatattatttttgacag 548
 Db 121 gagaaaggaaaattgtcagtagtcaatggaaccagatcggaataaccattttggacaa 180
 QY 549 gtggtacgcttgaatgcaactccggttcaagattgaaggacagaaagaatgcaactgc 608
 Db 181 gcagtcaggttgtatgtaactcaggctacaagattgaaggagatgaagaatgcattgt 240
 QY 609 tcataaaatggcctctggagcaatgaaaagccacagtggtgtgaaaattttctgcctgcc 668
 Db 241 tcagacgatggttttggagtaagaagaacacaaagtgtgtgaaatttcattgcataatcc 300
 QY 669 ccacgagttgaaaatgga 686
 Db 301 ccagatgttataaatgga 318

RESULT 3
 AAV02791
 ID AAV02791 standard; DNA; 581 BP.
 XX AC AAV02791;
 XX DT 27-APR-1998 (first entry)
 XX DE Clone PRRB9FH410 CFH related protein DNA fragment.
 XX KW Complement factor H; tumour associated antigen; renal cancer;
 KW urogenital cancer; medicament; modulator; ss.
 XX OS Synthetic.
 XX PN WO9738136-A1.
 XX PD 16-OCT-1997.
 XX PF 09-APR-1997; 97WO-US05710.
 XX PR 06-MAR-1997; 97US-0812481.
 PR 09-APR-1996; 96US-0015083.
 PR 09-APR-1996; 96US-0630048.
 PR 06-MAR-1997; 97US-0038614.
 XX PA (BARD-) BARD DIAGNOSTIC SCI INC.
 XX PI Enfield DL, Hass GM, Kinders RJ;
 XX DR WPI; 1997-512742/47.
 XX P-PSDB; AAW39155.
 XX PT Treating or screening for cancer, e.g. renal or urogenital cancer -
 PT by modulating or detecting tumour associated human complement factor
 PT H related antigen, or nucleic acid encoding it
 XX PS Example 6B; Fig 6A; 104pp; English.
 XX CC This partial cDNA sequence is present in clone PRRB9FH410 and encodes a
 CC complement factor H related protein with homology to a region of the
 CC human tumour-associated complement factor H (CFH). The detection of this
 CC protein and a CFH antigen can be used in screening or for the treatment
 CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.
 CC Agents that may modulate this antigen could be used in the manufacture of
 CC a medicament for the treatment of a tumour cell.
 XX SQ Sequence 581 BP; 198 A; 85 C; 139 G; 159 T; 0 other;

Query Match 25.1%; Score 217.2; DB 18; Length 581;
 Best Local Similarity 80.2%; Pred. No. 2.7e-52;
 Matches 255; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
 QY 369 acatgtgatgaagggtaccactattagtgtaaatgattaccgtgaatgtgatgcagat 428

Db 1 acatgaatgaggggtatcaattgctaggtgagattaataccgtgaatgtgacacagat 60
 QY 429 gggtagacaaatgatatctcaatctatgtgaagtgtgaagtgtccagtgacagaaactg 488
 Db 61 ggaatgacaaatgattctctatctatgtgaagtgtgaagtgttaccagtgacagaccca 120
 QY 489 ggaatggaagaattgtgagtggtgcagccgaaccagaccaggaatattattttggacag 548
 Db 121 gagaaaggaaaattgtcagtagtcaatggaaccagatcggaataaccattttggacaa 180
 QY 549 gtggtacgcttgaatgcaactccggttcaagattgaaggacagaaagaatgcaactgc 608
 Db 181 gcagtcaggttgtatgtaactcaggctacaagattgaaggagatgaagaatgcattgt 240
 QY 609 tcataaaatggcctctggagcaatgaaaagccacagtggtgtgaaaattttctgcctgcc 668
 Db 241 tcagacgatggttttggggtaaagaagaacacaaagtgtgtgaaatttcattgcataatcc 300
 QY 669 ccacgagttgaaaatgga 686
 Db 301 ccagatgttataaatgga 318

RESULT 4
 AAH75787
 ID AAH75787 standard; cDNA; 3696 BP.
 XX AC AAH75787;
 XX DT 16-OCT-2001 (first entry)
 XX DE Receptor 222 coding sequence.
 XX KW Receptor 222; cytostatic; anti-HIV; cancer; HIV infection; ss.
 XX OS Unidentified.
 XX FH Key Location/Qualifiers
 FT CDS 402..3342
 FT /*tag= a
 FT /product= "Receptor 222"
 XX CN1296952-A.
 XX PN 30-MAY-2001.
 XX PD 23-NOV-1999; 99CN-0124086.
 XX PR 23-NOV-1999; 99CN-0124086.
 XX PA (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.
 XX PI Mao Y, Xie Y;
 XX WPI; 2001-483894/53.
 XX P-PSDB; AAG66398.
 XX PT New polypeptide-complement receptor 222 for treating diseases, such as,
 PT cancer and human immunodeficiency virus infection -
 XX Claim 6; Pages 19-21 (Disclosure); 30pp; Chinese.
 XX CC The present sequence is the coding sequence for receptor 222. Receptor
 CC 222 and its coding sequence are useful in treating diseases e.g. cancer
 CC and HIV infection.
 XX SQ Sequence 3696 BP; 977 A; 837 C; 895 G; 987 T; 0 other;

Query Match 5.3%; Score 46.2; DB 22; Length 3696;
 Best Local Similarity 59.5%; Pred. No. 0.0071;
 Matches 78; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

RESULT	7	
ABA06548		
ID	ABA06548 standard; cDNA; 3804 BP.	
XX	ABA06548;	
XX	10-JAN-2002 (first entry)	
XX	Human cDNA SEQ ID NO: 214.	
XX	Human; gene therapy; neural disorder; immune system disorder;	
KW	muscular disorder; reproductive disorder; gastrointestinal disorder;	
KW	pulmonary disorder; cardiovascular disorder; renal disorder;	
XX	proliferative disorder; inflammation; ss.	
XX	Homo sapiens.	
XX	WO200154474-A2.	
PN	02-AUG-2001.	
XX	17-JAN-2001; 2001WO-US01349.	
PD	31-JAN-2000; 2000US-179065P.	
XX	04-FEB-2000; 2000US-180628P.	
XX	24-FEB-2000; 2000US-184664P.	
PR	02-MAR-2000; 2000US-186350P.	
PR	16-MAR-2000; 2000US-189874P.	
PR	17-MAR-2000; 2000US-190076P.	
PR	18-APR-2000; 2000US-196123P.	
PR	19-MAY-2000; 2000US-205515P.	
PR	07-JUN-2000; 2000US-209467P.	

OY 651 gaaattcttg 661
| | | | |
Db 1886 cacattcttg 1896

RESULT 9
ID AAS26868
XX AAS26868 standard; cDNA; 3804 BP.
AC AAS26868;
XX
XX
DT 07-NOV-2001 (first entry)
TX Human cDNA encoding a novel secreted protein, SEQ ID 60.
DE XX
DE Human; immunosuppressive; antiarthritic; ss; antirheumatic;
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
XX
PN WO20015441-A2.
PD
PD 02-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001WO-US01320.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216680.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:00:49 ; Search time 39.15 Seconds
(without alignments)
508.059 Million cell updates/sec

Title: US-09-316-163-9
Perfect score: 1163
Sequence: 1 EDNCNELPPRRNTEILTGSMW.....VEISCKSPDVINGSPISQKI 207
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1163	100.0	449	1	NBHHS
2	1163	100.0	1231	1	complement factor
3	822	70.7	1234	1	complement factor
4	499	42.9	669	2	factor H - bovine
5	368.5	31.7	1053	2	probable complement
6	275	23.6	360	2	complement control
7	262	22.5	597	1	C4BP alpha chain p
8	261	22.4	381	1	decay-accelerating
9	261	22.4	440	2	decay-accelerating
10	260	22.4	676	2	complement C3b rec
11	255	21.9	2489	2	E-selectin - pig
12	253.5	21.8	482	2	decay-accelerating
13	251	21.6	340	2	complement recepto
14	247	21.2	2014	2	apolipoprotein H h
15	246	21.0	263	1	C4b-binding protei
16	244	21.0	597	1	hypothetical prote
17	243.5	20.9	497	2	complement recepto
18	243	20.9	560	2	complement recepto
19	240.5	20.7	440	2	hypothetical prote
20	238.5	20.5	349	2	sperm CD46 - human
21	238.5	20.5	369	2	membrane cofactor
22	238.5	20.5	377	2	membrane cofactor
23	238.5	20.5	384	2	membrane cofactor
24	236	20.3	263	1	complement control
25	236	20.3	263	1	hypothetical prote
26	236	20.3	302	1	secretory compleme
27	236	20.3	360	1	membrane-bound com
28	235	20.2	462	2	B18L protein - var
29	235	20.2	462	2	complement C3b/C4b

30	230.5	19.8	610	2	A35046
31	229.5	19.7	362	2	JC5194
32	229.5	19.7	369	2	JC5138
33	228	19.6	469	1	NBMS4
34	227.5	19.6	768	2	A42755
35	227.5	19.6	830	2	A30359
36	222	19.1	768	2	I53821
37	218.5	18.8	612	2	B42755
38	216	18.6	558	2	S57953
39	216	18.6	977	2	I52657
40	214.5	18.4	551	2	I46709
41	214.5	18.4	610	1	I46001
42	212	18.2	1025	1	A43526
43	210.5	18.1	646	2	JN0473
44	210.5	18.1	1091	1	PL0009
45	210	18.1	372	1	A32375

ALIGNMENTS

RESULT 1

NBHHS
complement factor H precursor, short splice form [validated] - human
N:Alternate names: complement factor H-related protein; complement protein H
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 23-Feb-1996 #text_change 08-Dec-2000
C:Accession: S03013; B60238; A27877; A61103; A26505; S10479
R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.
Biochem. J. 249, 593-602, 1988
A:Title: The complete amino acid sequence of human complement factor H.
A:Reference number: S00254; MUID:88134059
A:Accession: S03013
A:Molecule type: mRNA
A:Residues: 1-449 <RIP>
A:Cross-references: EMBL:X07523; EMBL:Y00716; NID:932492; PIDN:CAA30403.1; PID:975807
A:Note: part of this sequence, including the amino end of the mature protein was conf
A:Note: 402-Tyr was also found
R:Estaller, C.; Schwaebler, W.; Dierich, M.; Weiss, E.H.
Eur. J. Immunol. 21, 799-802, 1991
A:Title: Human complement factor H: two factor H proteins are derived from alternativ
A:Reference number: A60238; MUID:91184292
A:Accession: B60238
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-33;434-449 <EST>
A:Note: Only portions of this 1.8 kilobase mRNA were sequenced
R:Schulz, T.F.; Schwaebler, W.; Stanley, K.K.; Weiss, E.; Dierich, M.P.
Eur. J. Immunol. 16, 1351-1355, 1986
A:Title: Human complement factor H: isolation of cDNA clones and partial cDNA sequenc
A:Reference number: A27877; MUID:87054207
A:Accession: A27877
A:Molecule type: mRNA
A:Residues: '1L',55-401,'Y',403-449 <SCH>
A:Cross-references: GB:X04697; NID:931991; PIDN:CAB41739.1; PID:94725976
A:Note: an additional nucleotide present within the codon for Glu-310 was thought to
R:Schwaebler, W.; Zwirner, J.; Schulz, T.F.; Linke, R.P.; Dierich, M.P.; Weiss, E.H.
Eur. J. Immunol. 17, 1485-1489, 1987
A:Title: Human complement factor H: expression of an additional truncated gene produc
A:Reference number: A61103; MUID:88055295
A:Accession: A61103
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 27-76 <SC2>
A:Note: this is a partial sequence of an alternatively spliced 1.8 kilobase mRNA that
R:Sim, R.B.; DiScipio, R.G.
Biochem. J. 205, 285-293, 1982
A:Title: Purification and structural studies on the complement-system control protein
A:Reference number: A26505; MUID:83048213
A:Accession: A26505
A:Molecule type: protein
A:Residues: 19-20,'Q',22-29,'V',31-33,'Q',35 <SIM>
R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P

Biochemistry 31, 3626-3634, 1992

A:Title: Solution structure of the fifth repeat of factor H: A second example of the complement factor H precursor, long splice form [validated] - human

A:Reference number: A44551; MUID:92232649

A:Keywords: complement factor H, short splice form #status experimental <NAT>

R:Krisstensen, T.; Wetsel, R.A.; Tack, B.F.

J. Immunol. 136, 3407-3411, 1986

A:Title: Structural analysis of human complement protein H: homology with C4b binding protein

A:Reference number: S10479; MUID:86169701

A:Accession: S10479

A:Molecule type: mRNA

A:Residues: 226-401, 'Y', 403-449 <KRI>

A:Cross-references: GB:M12383; NID:g180472; PIDN:AAA52013.1; PID:g180473

C:Comment: Factor H has also been found bound to cell membranes in an unknown manner. Hc

C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver.

C:Genetics: <HF1>

A:Gene: GDB:HF1; HF

A:Cross-references: GDB:120041; OMIM:134370

A:Map position: lq32-lq32

C:Genetics: <HF2>

A:Gene: GDB:HF2; HF

A:Cross-references: GDB:129095

A:Map position: lq32-lq32

A:Note: the correspondence between the two loci and the sequences indicated is unclear;

C:Function:

A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also increase alternative complement pathway

A:Pathway: complement alternate pathway

C:Superfamily: complement factor H; complement factor H repeat homology

C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-449/Product: complement factor H, short splice form #status experimental <NAT>

F:21-80/Domain: complement factor H repeat homology <FH01>

F:85-141/Domain: complement factor H repeat homology <FH02>

F:146-205/Domain: complement factor H repeat homology <FH03>

F:210-262/Domain: complement factor H repeat homology <FH04>

F:246-248/Region: cell attachment (R-G-D) motif

F:267-320/Domain: complement factor H repeat homology <FH05>

F:325-385/Domain: complement factor H repeat homology <FH06>

F:389-442/Domain: complement factor H repeat homology <FH07>

F:21-66, 52-80, 85-129, 114-141, 146-192, 178-205, 210-251, 237-262, 267-309, 294-320, 325-374, 375-401/Binding site: carbohydrate (Asn) (covalent) #status absent

F:217/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 100.0%; Score 1163; DB 1; Length 449;

Best Local Similarity 100.0%; Pred. No. 1.9e-90;

Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDCNELPRRNTTEILTGSWSDDQTPGCTQAIYKCRPCYSLGNVIMVCRKGWVALLNPLR 60

Db 19 EDCNELPRRNTTEILTGSWSDDQTPGCTQAIYKCRPCYSLGNVIMVCRKGWVALLNPLR 78

Qy 61 KCQRKPCGHPGDPFTFTLTGGNVFYGKAVYTCNEGQYLLGEINRYECDDTGWNTDI 120

Db 79 KCQRKPCGHPGDPFTFTLTGGNVFYGKAVYTCNEGQYLLGEINRYECDDTGWNTDI 138

Qy 121 PICEVVKCLPVTAPENKIVSSAMEPQREYHFGQAVFVCSNGYKIEGDEMHCSDDGF 180

Db 139 PICEVVKCLPVTAPENKIVSSAMEPQREYHFGQAVFVCSNGYKIEGDEMHCSDDGF 198

Qy 181 SKERPKCIVEISCKSPDVINGSPISOKI 207

Db 199 SKERPKCIVEISCKSPDVINGSPISOKI 225

RESULT 2

NBHUH

C:Species: Homo sapiens (man)

C:Date: 31-Dec-1993 #sequence.revision 31-Dec-1993 #text.change 08-Dec-2000

C:Accession: S00254; A60238; A61726; A61565; A26505; I72654; S66298

R:Ripoczne, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.

Biochem. J. 249, 593-602, 1988

A:Title: The complete amino acid sequence of human complement factor H.

[illegible]

Db 139 PLCEVVKCLPVTLENGRIVSGAAETDQYFQGVVRFECNSGFKTEGHKEHCSENGLW 198
QY 181 SREKPKVCVEISKSPDVINGSPISOK 206
Db 199 SNEKPRCVELCTPPRVENGDDGINKV 224

RESULT 4

S65551
factor H - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S65551
R:Soames, C.J.; Day, A.J.; Sim, R.B.
Biochem. J. 315, 523-531, 1996
A:Title: Prediction from sequence comparisons of residues of factor H involved in the in
A:Reference number: S65551; MUID:96202005
A:Accession: S65551
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-669 <SOA>
A:Cross-references: GB:X98697; NID:q1419423; PIDN:CAA67257.1; PID:q1419424
C:Superfamily: complement factor H; complement factor H repeat homology
F:55-114/Domain: complement factor H repeat homology <FHR1>
F:296-349/Domain: complement factor H repeat homology <FHR2>
F:355-412/Domain: complement factor H repeat homology <FHR3>
F:416-471/Domain: complement factor H repeat homology <FHR4>
F:476-530/Domain: complement factor H repeat homology <FHR5>
F:538-592/Domain: complement factor H repeat homology <FHR5>
F:599-651/Domain: complement factor H repeat homology <FHR6>

Query Match 42.9%; Score 499; DB 2; Length 669;
Best Local Similarity 64.9%; Pred. No. 2.5e-34;
Matches 85; Conservative 19; Mismatches 27; Indels 0; Gaps 0;

QY 76 GFTLTGGNVFEGYKAVYTCNEGYYQLLGEINRECDTGTGNTNDIPICEVVKCLPVTAPPE 135
Db 3 GSPHLAGNQFEGYKAVYTCNEGYYQLLGEINRECDTGTGNTNDIPICEVVKCLPVTPE 62
QY 136 NGKIVSSAMEPREYHFGQAVFVCSNGYKIEGDEMHCSDDGFSWKEKPKVCVEISKSP 195
Db 63 NCKIFSDALEPQDEYTYGQVQFECNSGYMLDGPQKHCSAGGVWSAETPKVCVEIFCKPP 122
QY 196 DVINGSPISOK 206
Db 123 VILNGQAVLPK 133

RESULT 5

S46199
Probable complement regulatory plasma protein SB1 - barred sand bass
C:Species: Paralabrax nebulifer
C:Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 18-Jul-2001
C:Accession: S46199; S77894
R:Dahmen, A.; Kaidoh, T.; Zipfel, P.F.; Gagli, I.
Biochem. J. 301, 391-397, 1994
A:Title: Cloning and characterization of a cDNA representing a putative complement-regul
A:Reference number: S46199; MUID:94318039
A:Accession: S46199
A:Molecule type: mRNA
A:Residues: 1-1053 <DAH1>
A:Cross-references: EMBL:L21703; NID:g639894; PIDN:AAA92556.1; PID:g639895
A:Experimental source: liver
A:Accession: S77894
A:Molecule type: protein
A:Residues: 526-532, 'X', 534-537, 809-817, 'X', 819-826 <DAH2>
C:Genetics:
A:Gene: SB1
C:Superfamily: complement factor H repeat homology
C:Keywords: glycoprotein
F:89-145/Domain: complement factor H repeat homology <FH01>

F:334-389/Domain: complement factor H repeat homology <FH02>
F:450-502/Domain: complement factor H repeat homology <FHR1>
F:569-624/Domain: complement factor H repeat homology <FH03>
F:682-738/Domain: complement factor H repeat homology <FH04>
F:743-802/Domain: complement factor H repeat homology <FHR>
F:935-989/Domain: complement factor H repeat homology <FH05>
F:993-1052/Domain: complement factor H repeat homology <FH06>

Query Match 31.7%; Score 368.5; DB 2; Length 1053;
Best Local Similarity 39.6%; Pred. No. 3.9e-23;
Matches 72; Conservative 25; Mismatches 76; Indels 9; Gaps 4;

QY 21 DQTYPEGTQAIYKCRPGYSLGNVIMVCRKGEWVALNPLRKCQKRRCPGHPGDPFGFTTL 80
Db 47 EASYPGGRQVRVCCNGYS--GFFKLVCVGGKWEETR--AKQPRSCGHPGDAQADFHL 102
QY 81 TGGNVEYGVKAVYTCNEGYYQLLGEINRECDTGTGNTNDIPICEVVKCLPVTAPENGKIV 140
Db 103 AEGNDFVFGSKVYTCQGYMSRINRYRCVAEGDGVVYVCGESQCCPLIHVDNNVQVI 162
QY 141 SSAMEPDRVHFQAVRFVCSNGYKI--EGDEMHCSDDGFSWKEKPKVCVEISKSPDVIN 199
Db 163 GG-----PEATPGNVVRFSCSKSRSEILDGSPELYCDERGDWSGVPKCKAITCAIPPIEN 218
QY 200 GS 201
Db 219 GN 220

RESULT 6

T42921
complement control protein homolog ccph - ateline herpesvirus 3 (strain 73)
C:Species: ateline herpesvirus 3
A:Variety: strain 73
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 04-Mar-2000
C:Accession: T42921
R:Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A:Description: Primary structure of the herpesvirus ateles genome.
A:Reference number: 222274
A:Accession: T42921
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-360 <ALB>
A:Cross-references: EMBL:AF083424; PIDN:AAC95530.1
A:Experimental source: strain 73
C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom

Query Match 23.6%; Score 275; DB 2; Length 360;
Best Local Similarity 37.1%; Pred. No. 9.4e-16;
Matches 78; Conservative 20; Mismatches 80; Indels 32; Gaps 14;

QY 8 PRNTEI-----LTGNSDQYTPGCTQAIYKCRPGYSLGNVI--MVCRKGEWVALNPL 59
Db 24 PKRNVYSLRYVNIT--SSGYPNGTTLQVTCRKGY--IGRQIQVTCVNGNWTVPN-- 77
QY 60 RKCQKRPCGHPGTPGTFTLTGGNVFEGYKAVYTCNEGYYQLLGEINRECDTGTGNTNDIPICEVVKCLPVTAPENGKIV 115
Db 78 -ECQRRRCSTPADLLNGWTVT--GNLY--YGSVITYTCNTGYQLLGSFT--SSCLLGPGRV 133
QY 116 -WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDRVHF--FQAVRFVCSNGYK--IEGDEE 171
Db 134 NNTPRPPEICETKCRPPPTIANGTHTNI-----KEYTYLDAVYTCNDETCLTLTGPS 188
QY 172 MHCSDDGFW-SKEKPKVCVEISKSPDVING 200
Db 189 KQCSSETGRVWPDEETKCEKVCCKIPQVANG 218

RESULT 7

S53711

C4BP alpha chain precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S53711
R:de Frutos, P.G.; Dahlbaeck, B.
Biochim. Biophys. Acta 1261, 285-289, 1995
A:Title: cDNA structure of rabbit C4b-binding protein alpha-chain. Preserved sequence md
A:Reference number: S53711; MUID:95226458
A:Accession: S53711
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-597 <DEF>
A:Cross-references: EMBL:Z35490
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
F:50-107/Domain: complement factor H repeat homology <FH1>
F:112-169/Domain: complement factor H repeat homology <FH2>
F:174-234/Domain: complement factor H repeat homology <FH3>
F:239-294/Domain: complement factor H repeat homology <FH4>
F:299-360/Domain: complement factor H repeat homology <FH5>
F:364-422/Domain: complement factor H repeat homology <FH6>
F:426-480/Domain: complement factor H repeat homology <FH7>
F:484-538/Domain: complement factor H repeat homology <FH8>

Query Match 22.5%; Score 262; DB 1; Length 597;
Best Local Similarity 32.2%; Pred. No. 2e-14;
Matches 69; Conservative 30; Mismatches 93; Indels 22; Gaps 10;

QY 2 DCNELPPRRNTEILTGSWSDDQTYPECTQAIYKCRPGYSLG-NVIMVCR-KGEWVALNPL 59
Db 49 DCGP-PPHLLFASSISELSENEVQTGILKTCRPGYTRNGLNPLILTKRPLGLW---SYD 104

QY 60 RKCQRPCHGPDTPGTTFLTGGNVFEGVAVYTCNGYQLLGEINRECDT-----G 115
Db 105 TFCVKRRCRNPGLPQGQVEK--TDFSGQSFSESGEYLIGSTT-SHCDIQEKGEV 161

QY 116 WTNDIPICVWVVKCLPVTAPENGKIVSSAMEPDREYH-FGOAVRFVCSNGYKIEGDEMH 174
Db 162 WSDLPKCEIVKCEPPNLIINGKHGG-----NEDHTYGSSTYSCNPRFSLLEASISC 217

QY 175 S-----DDGFWMSKEKPKVCBISKSPDVINGSPIS 204
Db 218 TVKNKTVGVWSPSPVCKEIIICSPNPVPHGKIIS 251

RESULT 8
B26359
decay-accelerating factor, GPI-anchored splice form precursor - human
N:Alternate names: CD55; DAF
C:Species: Homo sapiens (man)
C:Date: 05-Oct-1988 #sequence_revision 16-Aug-1996 #text_change 19-Jan-2001
C:Accession: B26359; A27666; A39101; I52594; I52564
R:Caras, I.W.; Davitz, M.A.; Rhee, L.; Weddell, G.; Martin Jr., D.W.; Nussenzweig, V.
Nature 325, 545-549, 1987
A:Title: Cloning of decay-accelerating factor suggests novel use of splicing to generate
A:Reference number: A26359; MUID:87115645
A:Accession: B26359
A:Molecule type: mRNA
A:Residues: 1-381 <CAR>
A:Cross-references: GB:M30142; NID:g181464; PIDN:AAA52168.1; PID:g181465
R:Medof, M.E.; Lublin, D.M.; Hollers, V.M.; Ayers, D.J.; Getty, R.R.; Leykam, J.F.; Atkin
Proc. Natl. Acad. Sci. U.S.A. 84, 2007-2011, 1987
A:Title: Cloning and characterization of cDNAs encoding the complete sequence of decay-a
A:Reference number: A27666; MUID:87175602
A:Accession: A27666
A:Molecule type: mRNA
A:Residues: 6-79, '81-84, '86-381 <MED>
A:Cross-references: GB:M15799; NID:g181462; PIDN:AAA52167.1; PID:g181463
R:Moran, P.; Raab, H.; Kohn, W.J.; Caras, I.W.
J. Biol. Chem. 266, 1250-1257, 1991
A:Title: Glycophospholipid membrane anchor attachment. Molecular analysis of the cleavag
A:Reference number: A39101; MUID:91093238
A:Accession: A39101

A:Molecule type: protein
A:Residues: 338-352 <MOR>
R:Tublin, D.M.; Mallinson, G.; Poole, J.; Reid, M.E.; Thompson, E.S.; Ferdman, B.R.;
Blood 84, 1276-1282, 1994
A:Title: Molecular basis of reduced or absent expression of decay-accelerating factor
A:Reference number: I52594; MUID:94325573
A:Accession: I52594
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 194-198, 'L', 200-209 <LUB>
A:Cross-references: GB:S72858; NID:g639599; PIDN:AAC60633.1; PID:g639600
A:Experimental source: individual KW, Cromer blood group phenotype Dr(a-)
A:Note: the single nucleotide difference in this allele, which changes Ser-199 to Leu
on (see reference I52564), and thus reduced DAF expression
R:Reid, M.E.; Mallinson, G.; Sim, R.B.; Poole, J.; Pausch, V.; Merry, A.H.; Liew, Y.W
Blood 78, 3291-3297, 1991
A:Title: Biochemical studies on red blood cells from a patient with the Inab phenotype
A:Reference number: I52564; MUID:92075980
A:Accession: I52564
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 190-193, 'QLCPVE' <RE2>
A:Cross-references: GB:S70688; NID:g240301; PIDN:AAB20576.1; PID:g240302
A:Experimental source: individual KW, Cromer blood group phenotype Dr(a-) (described
C:Comment: Cromer blood group system antigens reside on this protein.
C:Comment: For an alternative splice form, see PIR:A26359.
C:Genetics:
A:Gene: GDB:DAF
A:Cross-references: GDB:I19088; OMIM:125240
A:Map position: lq32-lq32
C:Function:
A:Description: protects tissues from damage by regulating complement activation on ce
C:Superfamily: decay-accelerating factor; complement factor H repeat homology
C:Keywords: alternative splicing; blocked carboxyl end; complement inhibitor; glycopr
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-353/Product: decay-accelerating factor 2 #status predicted <MAT>
F:36-94/Domain: complement factor H repeat homology <FH01>
F:98-158/Domain: complement factor H repeat homology <FH02>
F:163-220/Domain: complement factor H repeat homology <FH03>
F:225-283/Domain: complement factor H repeat homology <FH04>
F:354-381/Domain: carboxyl-terminal propeptide #status predicted <PRO>
F:95/Binding site: carboxylate (Asn) (covalent) #status predicted
F:353/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature f

Query Match 22.4%; Score 261; DB 1; Length 381;
Best Local Similarity 29.9%; Pred. No. 1.5e-14;
Matches 70; Conservative 42; Mismatches 84; Indels 38; Gaps 12;

QY 1 EDCN----ELPPRRNTEILTGSWSDDQTY-PEGTQAIYKCRPGYR---SLGNVIMVCRKEW 53
Db 92 EFCNRCEVPTRLNSASLQPYITQNYFPVGTVEVECRPGYRREPSLSPLKTLCLQNLKW 151

QY 54 VALNPLRKCQRPCHGPDTPGTTFLTGGNVFEGVAVYTCNGYQLLGEINRECDT 113
Db 152 --STAVEFCCKKSCPNPGEIRNGQIDVPGGILF--GATISFCSCNTGYKLFGSTS-SFCLI 206

QY 114 DG-----WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDRE-YHFGOAVRFVCSNGYKIEG 168
Db 207 SSSVQVMSDPLDECRYYC-----PAPPQIDNGIIGENDHYGVSIVTAKNGKFTMIG 261

QY 169 DEEMHC---SDGFWMSKEKPKC-----VEISCKSPDVIN-----GSPISOK 206
Db 262 EHSIVCTVNNDEGWSGPPPECRGKSLTSKVPTTVQKPTTVNVPTEVSPTSOK 315

RESULT 9
A26359
decay-accelerating factor, splice form 1 precursor - human
N:Alternate names: decay-accelerating factor 5; decay-accelerating factor secreted fo
C:Species: Homo sapiens (man)
C:Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 02-Jun-2000
C:Accession: A26359; A39702; SI6187; S23138; A27258

F:238-293/Domain:	complement factor	H repeat	homology <FH04>
F:297-353/Domain:	complement factor	H repeat	homology <FH05>
F:358-416/Domain:	complement factor	H repeat	homology <FH06>
F:421-487/Domain:	complement factor	H repeat	homology <FH07>
F:433-449/Domain:	complement factor	H repeat	homology <FH08>
F:454-611/Domain:	complement factor	H repeat	homology <FH09>
F:616-682/Domain:	complement factor	H repeat	homology <FH10>
F:688-743/Domain:	complement factor	H repeat	homology <FH11>
F:747-803/Domain:	complement factor	H repeat	homology <FH12>
F:808-866/Domain:	complement factor	H repeat	homology <FH13>
F:871-937/Domain:	complement factor	H repeat	homology <FH14>
F:943-999/Domain:	complement factor	H repeat	homology <FH15>
F:1004-1061/Domain:	complement factor	H repeat	homology <FH16>
F:1066-1132/Domain:	complement factor	H repeat	homology <FH17>
F:1138-1193/Domain:	complement factor	H repeat	homology <FH18>
F:1197-1253/Domain:	complement factor	H repeat	homology <FH19>
F:1258-1316/Domain:	complement factor	H repeat	homology <FH20>
F:1321-1387/Domain:	complement factor	H repeat	homology <FH21>
F:1393-1449/Domain:	complement factor	H repeat	homology <FH22>
F:1454-1511/Domain:	complement factor	H repeat	homology <FH23>
F:1516-1582/Domain:	complement factor	H repeat	homology <FH24>
F:1588-1643/Domain:	complement factor	H repeat	homology <FH25>
F:1647-1703/Domain:	complement factor	H repeat	homology <FH26>
F:1708-1766/Domain:	complement factor	H repeat	homology <FH27>
F:1771-1837/Domain:	complement factor	H repeat	homology <FH28>
F:1846-1902/Domain:	complement factor	H repeat	homology <FH29>
F:1967-1964/Domain:	complement factor	H repeat	homology <FH30>
F:1969-2035/Domain:	complement factor	H repeat	homology <FH31>
F:2041-2096/Domain:	complement factor	H repeat	homology <FH32>
F:2100-2156/Domain:	complement factor	H repeat	homology <FH33>
F:2161-2219/Domain:	complement factor	H repeat	homology <FH34>
F:2224-2290/Domain:	complement factor	H repeat	homology <FH35>
F:2298-2354/Domain:	complement factor	H repeat	homology <FH36>
F:2359-2415/Domain:	complement factor	H repeat	homology <FH37>

Query Match 21.9%; Score 255; DB 2; Length 2489;
 Best Local Similarity 30.3%; Pred. No. 3.6e-13;
 Matches 67; Conservative 39; Mismatches 83; Indels 32; Gaps

> 3 CNE---LPPRRNTEILTGSWSDQTYPEGTOAIYKCRPGYSLGNVMVCKRGEWALNPL 59
 || | | | | | | | | | | | | | | | | : | : | : | : | : | : | : | : | :
 Db 43 CNAPEWLFPARTNLT---DEFEFPIGTVLTNCEPCGYSGRPFSICLKNSVTGAKD- 97

> 60 RKQKRCPGHGPTPFTFTLLTGNNVFVEYGVAVYTCNEGYYQLLGNEYREC----DTDG 115
 : : : : : | | | | | | | | : : : : : | : | : | : | : | : | : | : | :
 Db 98 -RCRRKSCRNPDPVNGMWHIVKG--IQFGSQIKYSCTKYRLIGS--SSATCIISGDTVI 153

> 116 WTNDIPICEVVKC-LPTAPENKIYSSAMEPDREYHFGVAQVFCVNSG-----VKIEG 168
 | | | | | : | | | | | | | : | : | : | : | : | : | : | : | : | : | :
 Db 154 WDNETPICDRIPCGLPTTI-TNGDFISTNRK---NFHYGSVVYTRCNPCSGGRKRVFLVG 209

> 169 DEEMHC---SDD--GFWSKEPKKC-VETSCSKSPDVINGSPITS 204
 : : : | | | | | | | | | | | | | | | | | | | | | | : | : | : | : | : | :
 Db 210 EPSYICTSNDDQVINGSGPAQCIIIPNKCTPPNVENGILVS 250

RESULT 12

JC5092

E-selectin - pig

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 19-May-2000

R:Accession: JC5092

R:Winkler, H.; Brostjan, C.; Csizmadia, V.; Natarajan, G.; Anrather, J.; B

Gene 176, 67-72, 1996

A:Title: The intron-exon structure of the porcine E-selectin-encoding gene

A:Reference number: JC5092; MOID:97075911

A:Contents: endothelial cells

A:Accession: JC5092

A:Molecule type: DNA

A:Residues: 1-482 <WIN>

A:Cross-references: GB.U37521; NID:g1052974; PIDN:AC4680.1; PID:g1052975

A:Comments: This sequence was derived from a cDNA library constructed from

C:Date: 31-Dec-1989 #sequence_revision 30-Jun-1990 #text_change 22-Jun-1999
C:Accession: A31005; B42504
R:Kotwal, G.J.; Moss, B.
Nature 335, 176-178, 1988
A:Title: Vaccinia virus encodes a secretory polypeptide structurally related to complement
A:Reference number: A31005; MUID:88318974
A:Accession: A31005
A:Molecule type: DNA
A:Residues: 1-263 <KOT>
A:Cross-references: GB:X13166; NID:g60690; PIDN:CAA31564.1; PID:g60691
A:Experimental source: strain WR
R:Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 517-563, 1990
A:Title: Appendix to "The complete DNA sequence of vaccinia virus".
A:Reference number: A42501
A:Accession: B42504
A:Molecule type: DNA
A:Residues: 1-263 <GOE>
A:Cross-references: GB:M35027; NID:g335317; PIDN:AAA47997.1; PID:g335345
A:Experimental source: strain Copenhagen
R:Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 247-266, 1990
A:Title: The complete DNA sequence of vaccinia virus.
A:Reference number: A42531; MUID:91021027
A:Contents: annotation; possible protein-coding frames
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: herpesvirus complement control protein; complement factor H repeat homolog
C:Keywords: duplication; extracellular protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-263/Product: C4b-binding protein homolog #status predicted <MAT>
F:21-81/Domain: complement factor H repeat homology <FH1>
F:86-143/Domain: complement factor H repeat homology <FH2>
F:148-201/Domain: complement factor H repeat homology <FH3>
F:206-261/Domain: complement factor H repeat homology <FH4>

Query Match 21.2%; Score 246; DB 1; Length 263;
Best Local Similarity 33.7%; Pred. No. 1.9e-13;
Matches 63; Conservative 23; Mismatches 83; Indels 18; Gaps 9;
Qy 20 SDQTYPEGTQAIYKCRPGYR--SLGNVIMVCRKGEWALNPLRKCQKRCGHPGDPFGT 77
Db : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
40 ANANYNIGDTIEYLCPLGPKYKQKMGPIYAKCTGTGTLFN---QCIKRRCPSPRIDNGQ 96
Qy 78 FTLTGGNVFEYGVAVYTCNKGYYOLLGEI-NVRECDTDS---WTNDIPICEVVKCLPVTA 133
Db : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
97 LDIGG---VDFGSSITYSCNSGYHLIGESKSYCELGSTGSMWNPEADICESVKCQSPPS 153
Qy 134 PENGKIVSSAMEPDREYHFGQAVRFVNCNSGYKIEGDEEMHGSDDGFWSKEKPKVCVEISCK 193
Db : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
154 ISNGR--HNGYED--FYTDGSVVTVSCNSGYSLIGNSGVLCSS-GGEWS-DPPTCOIVKCP 207
Qy 194 SPDVING 200
Db : : | |
208 HPTISNG 214

Search completed: August 29, 2002, 15:02:15
Job time: 86 sec

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	1163	100.0	1231	1	CFAH_HUMAN	P08603 homo sapien
2	822	70.7	1234	1	CFAH_MOUSE	P08609 mus musculus
3	523.5	45.0	685	1	CFAH_BOVIN	Q28085 bos taurus
4	261	22.4	381	1	CR1_HUMAN	P08174 homo sapien
5	255	21.9	2039	1	CR1_HUMAN	P17927 homo sapien
6	253.5	21.8	484	1	LEM2_PIG	P08110 sus scrofa
7	251	21.6	340	1	DAF_PONPY	P49457 pongo pygma
8	246.5	21.2	507	1	DAF_CAVPO	P60401 cavia porce
9	246	21.2	263	1	VCP_VACCV	P10998 vaccinia vi
10	244	21.0	597	1	C4BP_HUMAN	P04003 homo sapien
11	238.5	20.5	377	1	MCP_HUMAN	P15529 homo sapien
12	236	20.3	360	1	CCPH_HSVSA	Q01016 herpesvirus
13	231.5	19.9	390	1	DAF1_MOUSE	Q61475 mus musculus
14	230.5	19.8	610	1	LEM2_HUMAN	P16581 homo sapien
15	228	19.6	469	1	C4BP_MOUSE	P08607 mus musculus
16	227.5	19.6	768	1	LEM3_MOUSE	Q01102 mus musculus
17	227.5	19.6	830	1	LEM3_HUMAN	P16109 homo sapien
18	226	19.4	549	1	LEM2_RAT	P08105 rattus norv
19	222.5	19.1	407	1	DAF2_MOUSE	P08105 rattus norv
20	222	19.1	768	1	LEM3_RAT	P08106 rattus norv
21	220	18.9	611	1	LEM2_CANFA	P33730 canis fami
22	218.5	18.8	612	1	LEM2_MOUSE	Q00690 mus musculus
23	216	18.6	558	1	C4BP_RAT	Q63514 rattus norv
24	214.5	18.4	551	1	LEM2_RABIT	P27113 ryctcolagus
25	214.5	18.4	610	1	C4BP_BOVIN	Q28065 bos taurus
26	214	18.4	769	1	LEM3_SHEEP	P08109 ovis aries
27	212	18.2	1025	1	CR2_MOUSE	P19070 mus musculus
28	210.5	18.1	646	1	LEM3_BOVIN	P42201 bos taurus
29	210.5	18.1	1033	1	CR2_HUMAN	P20023 homo sapien
30	210	18.1	372	1	LEM1_MOUSE	P18337 mus musculus
31	205.5	17.7	485	1	LEM2_BOVIN	P08107 bos taurus
32	205	17.6	345	1	APHO_BOVIN	P17690 bos taurus
33	202	17.4	958	1	HIG_DROME	O09101 drosophila

RA	Vik D.P., Williams S.A.;
RL	Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RL	[8]
RP	SEQUENCE OF 1-9 FROM N.A.
RA	Dominguez O.;
RA	Thesis (1993), Hospital Trias I Pujol, Spain.
RL	[9]
RP	STRUCTURE BY NMR OF 927-985 (SUSHI 16).
RX	MEDLINE-91278097; PubMed=1829116;
RA	Norman D.G., Barlow P.N., Baron M., Day A.J., Sim B., Campbell I.D.;
RT	"three-dimensional structure of a complement control protein module
RT	in solution.";
J. Mol. Biol.	219:717-725(1991).
RL	[10]
RP	STRUCTURE BY NMR OF 264-322 (SUSHI 5).
RX	MEDLINE-9232649; PubMed=1533152;
RA	Barlow P.N., Norman D.G., Steinkasserer A., Horne T.J., Pearce J.,
RA	Driscoll P.C., Sim B., Campbell I.D.;
RT	"solution structure of the fifth repeat of factor H: a second example
RT	of the complement control protein module.";
Biochemistry	31:3626-3634(1992).
RL	[11]
RP	STRUCTURE BY NMR OF 866-985 (SUSHS 15 AND 16).
RX	MEDLINE-93323119; PubMed=8331663;
RA	Barlow P.N., Steinkasserer A., Norman D.G., Kieffer B., Wiles A.P.,
RA	Sim B., Campbell I.D.;
RT	"Solution structure of a pair of complement modules by nuclear
RT	magnetic resonance.";
J. Mol. Biol.	232:268-284(1993).
CC	-1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF
CC	C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE
CC	C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3B)NBB COMPLEX (C5
CC	CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.
CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC	PRODUCED BY ALTERNATIVE SPLICING.
CC	-1- SIMILARITY: CONTAINS 20 SUSHI (SCR) DOMAINS
CC	-1- CAUTION: REF. 2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION
CC	341 ONWARD DUE TO A FRAMESHIFT.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outpost -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to licenset@isb-sib.ch).
CC	-----
EMBL;	Y00716; CAA68704.1;
DR	EMBL; X04697; CAB41739.1; ALT FRAME.
DR	EMBL; M12523; CAA30403.1;
DR	EMBL; M12383; AAA52013.1;
DR	EMBL; AL049744; CAB70598.1;
DR	EMBL; M65294; AAA35948.1;
DR	EMBL; U56979; AAB01987.1;
DR	EMBL; Z29665; CAA82763.1;
DR	PIR; S00254; NBHUH.
DR	PIR; S03013; S03013.
DR	PDB; 1HCC; 15-APR-92.
DR	PDB; 1HPF; 15-JUL-93.
DR	PDB; 1HFI; 15-JUL-93.
DR	MIM; I34370;
DR	InterPro: IPR000436; Sushi_SCR_CCP.
DR	Pfam: PF00084; sushi; 20.
DR	SMART: SM00032; CCP; 20.
KW	Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;
KW	Signal; 3D-structure; Polymorphism; Alternative splicing.
FT	SIGNAL 1 18
FT	CHAIN 19 1231
FT	DOMAIN 20 81
FT	DOMAIN 84 142
FT	DOMAIN 145 206
FT	DOMAIN 209 263
FT	DOMAIN 266 321
FT	COMPLEMENT FACTOR H.
FT	SUSHI 1.
FT	SUSHI 2.
FT	SUSHI 3.
FT	SUSHI 4.
FT	SUSHI 5.
FT	STRAND
FT	876
FT	CONFLICT
FT	CONFLICT
FT	CONFLICT
FT	T -> Q (IN REF. 3).
FT	T -> V (IN REF. 3).
FT	RP -> IL (IN REF. 2).
FT	RP -> Y.
FT	/Ftid-VAR_001979.
FT	C -> Q (IN REF. 3).
FT	T -> V (IN REF. 3).
FT	T -> Q (IN REF. 3).
FT	RP -> IL (IN REF. 2).
FT	RP -> Y.
FT	/Ftid-VAR_001979.
FT	C -> Q (IN REF. 3).
FT	T -> V (IN REF. 3).
FT	T -> Q (IN REF. 3).
FT	RP -> IL (IN REF. 2).
FT	RP -> Y.
FT	/Ftid-VAR_001979.
FT	C -> Q (IN REF. 3).
FT	T -> V (IN REF. 3).
FT	T -> Q (IN REF. 3).
FT	RP -> IL (IN REF. 2).
FT	RP -> Y.
FT	/Ftid-VAR_001979.
FT	C -> Q (IN REF. 3).
FT	T -> V (IN REF. 3).
FT	T -> Q (IN REF. 3).
FT	RP -> IL (IN REF. 2).
FT	RP -> Y.
FT	/Ftid-VAR_001979.
FT	C -> Q (IN REF. 3).
FT	T -> V (IN REF. 3).
FT	T -> Q (IN REF. 3).
FT	RP -> IL (IN REF. 2).
FT	RP -> Y.
FT	/Ftid-VAR_001979.
FT	C -> Q (IN REF. 3).
FT	T -> V (IN REF. 3).
FT	T -> Q (IN REF. 3).
FT	RP -> IL (IN REF. 2).
FT	RP -> Y.
FT	/Ftid-VAR_001979.
FT	C -> Q (IN REF. 3).
FT	T -> V (IN REF. 3).
FT	T -> Q (IN REF. 3).
FT	RP -> IL (IN REF. 2).
FT	RP -> Y.
FT	/Ftid-VAR_001979.
FT	C -> Q (IN REF. 3).
FT	T -> V (IN REF. 3).
FT	T -> Q (IN REF. 3).
FT	RP -> IL (IN REF. 2).
FT	RP -> Y.
FT	/Ftid-VAR_001979.
FT	C -> Q (IN REF. 3).
FT	T -> V (IN REF. 3).
FT	T -> Q (IN REF. 3).
FT	RP -> IL (IN REF. 2).
FT	RP -> Y.
FT	/Ftid-VAR_001979.
FT	C -> Q (IN REF. 3).
FT	T -> V (IN REF. 3).
FT	T -> Q (IN REF. 3).
FT	RP -> IL (IN REF. 2).
FT	RP -> Y.
FT	

Query Match 100.0%; Score 1163; DB 1; Length 1231;
Best Local Similarity 100.0%; Pred. No. 2.3e-95;
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDCNLPNRRNTTEILTGSMSDQTPGTOAIYKRCGYRSLGNVIMVCRKGEWALNPLR 60
Db 19 EDCNLPNRRNTTEILTGSMSDQTPGTOAIYKRCGYRSLGNVIMVCRKGEWALNPLR 78
QY 61 KCQKPCGHPGDTPTFTLTGNGNFEYGVKAVYTCNEGYQLLGEINRECDTDGWTNDI 120
Db 79 KCQKPCGHPGDTPTFTLTGNGNFEYGVKAVYTCNEGYQLLGEINRECDTDGWTNDI 138
QY 121 PICEVVKCLPVTAPENKIVSSAMEPDRYHFGQAVRFVNCNGYKIEGDEEHKCSDDGEW 180
Db 139 PICEVVKCLPVTAPENKIVSSAMEPDRYHFGQAVRFVNCNGYKIEGDEEHKCSDDGEW 198
QY 181 SKEKPKCIVEISCKSPDVINGSPISQKI 207
Db 199 SKEKPKCIVEISCKSPDVINGSPISQKI 225

RESULT 2
ID CFAH_MOUSE STANDARD; PRT: 1234 AA.
AC P06909;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Complement factor H precursor (Protein beta-1-H).
GN HF1 OR CFH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86233353; PubMed=2940596;
RA Kristensen T., Tack B.F.;
RT "Murine protein H is comprised of 20 repeating units, 61 amino acids
in length."
RL Proc. Natl. Acad. Sci. U.S.A. 83:3963-3967(1986).
RN [2]
RP SEQUENCE OF 1-19 FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=90148935; PubMed=2533512;
RA Munoz-Canoves P., Tack B.F., Vik D.P.;
RT "Analysis of complement factor H mRNA expression: dexamethasone and
IFN-gamma increase the level of H in L cells."
RL Biochemistry 28:9891-9897(1989).
RN [3]
RP SEQUENCE OF 1-18 FROM N.A.
RX MEDLINE=90111033; PubMed=2136885;
RA Natsume-Sakai S., Nonaka M., Harada Y.N., Shreffler D.C.,
RA Moriwaki K.;
RT "Demonstration of an unusual allelic variation of mouse factor H by
the complete cDNA sequence of the H.2 allotype."
RL J. Immunol. 144:358-362(1990).
CC -1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF
C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE
C3BB COMPLEX (C3 CONVERTASE) AND THE (C3B)NBB COMPLEX (C5
CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.
CC -1- POLYMORPHISM: TWO CODOMINANT ALLELES OF FACTOR H ARE PRESENT IN
MICE.
CC -1- SIMILARITY: CONTAINS 20 SUSHI (SCR) DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC
DR EMBL; M12660; AAA37759.1; -
DR EMBL; J02891; AAA37795.1; -
DR EMBL; M31979; AAA37762.1; -
DR PIR; A26154; NBMSH.
DR HSP; P08603; LHFI.
DR MGD; MGI:88385; Cfh.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 20.
DR SMART; SM00032; CCP; 20.
KW Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;
KW Signal.
FT SIGNAL 1 18 BY SIMILARITY.
FT CHAIN 19 1234 COMPLEMENT FACTOR H.
FT DOMAIN 20 81 SUSHI 1.
FT DOMAIN 84 142 SUSHI 2.
FT DOMAIN 145 206 SUSHI 3.
FT DOMAIN 209 263 SUSHI 4.
FT DOMAIN 266 321 SUSHI 5.
FT DOMAIN 324 386 SUSHI 6.
FT DOMAIN 388 443 SUSHI 7.
FT DOMAIN 447 506 SUSHI 8.
FT DOMAIN 508 565 SUSHI 9.
FT DOMAIN 568 623 SUSHI 10.
FT DOMAIN 628 684 SUSHI 11.
FT DOMAIN 689 744 SUSHI 12.
FT DOMAIN 751 803 SUSHI 13.
FT DOMAIN 807 862 SUSHI 14.
FT DOMAIN 866 932 SUSHI 15.
FT DOMAIN 935 990 SUSHI 16.
FT DOMAIN 993 1049 SUSHI 17.
FT DOMAIN 1052 1108 SUSHI 18.
FT DOMAIN 1113 1169 SUSHI 19.
FT DOMAIN 1171 1234 SUSHI 20.
FT DISULFID 21 66 BY SIMILARITY.
FT DISULFID 52 80 BY SIMILARITY.
FT DISULFID 85 129 BY SIMILARITY.
FT DISULFID 114 141 BY SIMILARITY.
FT DISULFID 146 192 BY SIMILARITY.
FT DISULFID 178 205 BY SIMILARITY.
FT DISULFID 210 251 BY SIMILARITY.
FT DISULFID 237 262 BY SIMILARITY.
FT DISULFID 267 309 BY SIMILARITY.
FT DISULFID 294 320 BY SIMILARITY.
FT DISULFID 325 374 BY SIMILARITY.
FT DISULFID 357 385 BY SIMILARITY.
FT DISULFID 389 431 BY SIMILARITY.
FT DISULFID 416 442 BY SIMILARITY.
FT DISULFID 448 494 BY SIMILARITY.
FT DISULFID 477 505 BY SIMILARITY.
FT DISULFID 509 553 BY SIMILARITY.
FT DISULFID 536 564 BY SIMILARITY.
FT DISULFID 569 610 BY SIMILARITY.
FT DISULFID 597 622 BY SIMILARITY.
FT DISULFID 629 672 BY SIMILARITY.
FT DISULFID 658 683 BY SIMILARITY.
FT DISULFID 690 732 BY SIMILARITY.
FT DISULFID 718 743 BY SIMILARITY.
FT DISULFID 752 791 BY SIMILARITY.
FT DISULFID 780 802 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 861 BY SIMILARITY.
FT DISULFID 867 920 BY SIMILARITY.
FT DISULFID 906 931 BY SIMILARITY.
FT DISULFID 936 978 BY SIMILARITY.
FT DISULFID 964 989 BY SIMILARITY.
FT DISULFID 994 1037 BY SIMILARITY.
FT DISULFID 1023 1048 BY SIMILARITY.
FT DISULFID 1053 1096 BY SIMILARITY.
FT DISULFID 1082 1107 BY SIMILARITY.
FT DISULFID 1114 1157 BY SIMILARITY.
FT DISULFID 1143 1168 BY SIMILARITY.
FT DISULFID 1172 1223 BY SIMILARITY.

```

CC      EMBL; X98697; CAA67257.1; -.
DR      HISSP; P10998; lVVD.
DR      InterPro; IPR000436; Sush1_SCR_CCP.
DR      Pfam; PF00084; sush; 11.
DR      SMART; SM00032; CCP; 11.

```

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;
 [1] SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RP MEDLINE=87115845; PubMed=2433596;
 RA Caras I.W., Davitz M.A., Rhee L., Weddell G., Martin D.W. Jr.,
 RA Nussenzweig V.;
 RT "Cloning of decay-accelerating factor suggests novel use of splicing
 RT to generate two proteins.";
 RL Nature 325:545-549(1987).
 [2] SEQUENCE FROM N.A.
 RP STRAUSBERG R.;
 RA TISSUE=Cervix;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 [3] SEQUENCE OF 6-381 FROM N.A. (ISOFORM 2).
 RX MEDLINE=87173602; PubMed=2436222;
 RA Medof M.E., Lublin D.M., Holers V.M., Ayers D.J., Getty R.R.,
 RA Leykam J.F., Atkinson J.P., Tykocinski M.L.;
 RT "Cloning and characterization of cDNAs encoding the complete sequence
 RT of decay-accelerating factor of human complement.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:2007-2011(1987).
 [4] SEQUENCE OF 35-381 FROM N.A. (ISOFORM 2).
 RP TISSUE=Hippocampus;
 RC Kumar V.B., Hyung C., Nakra R., Walters M., Sasser T., Bernardo A.;
 RA "Decay-accelerating factor (DAF; CD 55) in the brain of Alzheimer's
 RT disease patients.";
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 [5] SEQUENCE OF 1-100 FROM N.A.
 RP MEDLINE=91271256; PubMed=1711208;
 RA Ewulonu U.K., Ravi L., Medof M.E.;
 RT "Characterization of the decay-accelerating factor gene promoter
 RT region.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4675-4679(1991).
 [6] SEQUENCE OF 35-46.
 RP TISSUE=Urine;
 RC MEDLINE=91291869; PubMed=1712233;
 RA Nakano Y., Sugita Y., Ishikawa Y., Choi N.-H., Tobe T., Tomita M.;
 RT "Isolation of two forms of decay-accelerating factor (DAF) from human
 RT urine.";
 RL Biochim. Biophys. Acta 1074:326-330(1991).
 [7] GPI-ANCHOR.
 RP MEDLINE=91093238; PubMed=1824659;
 RA Moran P., Raab H., Kohr W.J., Caras I.W.;
 RT "Glycophospholipid membrane anchor attachment. Molecular analysis of
 RT the cleavage/attachment site.";
 RL J. Biol. Chem. 266:1250-1257(1991).
 [8] DISULFIDE BONDS IN SUSHI DOMAINS.
 RP MEDLINE=92305034; PubMed=1377029;
 RA Nakano Y., Sumida K., Kikuta N., Miura N.-H., Tobe T., Tomita M.;
 RT "Complete determination of disulfide bonds localized within the short
 RT consensus repeat units of decay accelerating factor (CD55 antigen).";
 RL Biochim. Biophys. Acta 1116:235-240(1992).
 [9] FUNCTION AS A ECHOVIRUS RECEPTOR.
 RP MEDLINE=95045399; PubMed=7525274;
 RA Ward T., Pipkin P.A., Clarkson N.A., Stone D.M., Minor P.D.,
 RA Almond J.W.;
 RT "Decay-accelerating factor CD55 is identified as the receptor for
 RT echovirus 7 using CELFICS, a rapid immuno-focal cloning method.";
 RL EMBO J. 13:5070-5074(1994).
 [10] VARIANT BLOOD GROUP DR(A-).
 RP MEDLINE=94325573; PubMed=7519480;
 RA Lublin D.M., Mallinson G., Poole J., Reid M.E., Thompson E.S.,
 RA Ferdman B.R., Telen M.J., Anstee D.J., Tanner M.J.A.;
 RT "Molecular basis of reduced or absent expression of
 RT decay-accelerating factor in Cromer blood group phenotypes.";

Blood 84:1276-1282(1994).
 -|- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT
 CC CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT
 CC C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.
 CC INTERFERS WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND
 CC FACTOR B TO ENZYMATICALLY ACTIVE C2A AND BB AND THEREBY PREVENTS
 CC THE FORMATION OF C4B2A AND C3BBB, THE AMPLIFICATION CONVERTASES OF
 CC THE COMPLEMENT CASCADE.
 -|- FUNCTION: ALSO ACTS AS THE RECEPTOR FOR ECHOVIRUS 7 AND RELATED
 CC VIRUSES (ECHOVIRUSES 13, 21, 29 AND 33).
 -|- SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT
 CC HOMODIMER (MINOR FORM).
 -|- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
 -|- TISSUE SPECIFICITY: EXPRESSED ON THE PLASMA MEMBRANES OF ALL CELL
 CC TYPES THAT ARE IN INTIMATE CONTACT WITH PLASMA COMPONENT
 CC PROTEINS. IT IS ALSO FOUND ON THE SURFACES OF EPITHELIAL CELLS
 CC Lining EXTRACELLULAR COMPARTMENTS, AND VARIANTS OF THE MOLECULE
 CC ARE PRESENT IN BODY FLUIDS AND IN EXTRACELLULAR MATRIX.
 -|- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
 CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
 CC ACTIVE SITE ON SCR3 (BY SIMILARITY).
 -|- PTM: THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.
 -|- POLYMORPHISM: DAF IS RESPONSIBLE FOR THE CROMER BLOOD GROUP
 CC SYSTEM. IT CONSISTS OF AT LEAST SEVEN HIGH-INCIDENCE (CR(A),
 CC TC(A), DR(A), ES(A), WES(B), UMC, AND IFC) AND LOW-INCIDENCE
 CC (TC(B), TC(C), AND WES(A)) ANTIGENS THAT RESIDE ON DAF. IN THE
 CC CROMER PHENOTYPES DR(A-) AND INAB THERE IS REDUCED OR ABSENT
 CC EXPRESSION OF DAF, RESPECTIVELY. IN THE CASE OF THE DR(A-)
 CC PHENOTYPE, A SINGLE NUCLEOTIDE SUBSTITUTION WITHIN EXON 5 ACCOUNTS
 CC FOR TWO CHANGES: A SIMPLE AMINO ACID SUBSTITUTION THAT IS THE
 CC BASIS OF THE ANTIGENIC VARIATION, AND AN ALTERNATIVE SPLICING
 CC EVENT THAT UNDERLIES THE DECREASED EXPRESSION OF DAF IN THIS
 CC PHENOTYPE.
 -|- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
 -|- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
 CC (RCA) FAMILY.
 -|- DATABASE: NAME=PROW; NOTE=CD guide CD55 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd55.htm".

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; M31516; AAA52169.1; -;
 DR EMBL; M30142; AAA52168.1; -;
 DR EMBL; BC001288; AAH01288.1; -;
 DR EMBL; M15799; AAA52167.1; -;
 DR EMBL; U88576; ABA48622.1; -;
 DR EMBL; M64553; AAA52170.1; -;
 DR EMBL; M64356; AAA52170.1; JOINED.
 DR EMBL; S72858; AAC60633.1; -;
 DR PIR; B26359; B26359.
 DR PIR; A26359; A26359.
 DR PIR; S16187; S16187.
 DR PIR; A39101; A39101.
 DR PIR; S23138; S23138.
 DR HSSP; P08603; 1HCC.
 DR MIN; 125240; -;
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00084; sushi; 4.
 DR SMART; SM00032; CCP; 4.
 KW Complement pathway; Plasma; Glycoprotein; Membrane; Repeat;
 KW Alternative splicing; GPI-anchor; Signal; Sushi; Polymorphism;
 KW Blood group antigen.
 FT SIGNAL 1 34
 FT CHAIN 35 353 COMPLEMENT DECAY-ACCELERATING FACTOR.

```

RX MEDLINE=89035992; PubMed=2972794;
RA Klickstein L.B., Bartow T.J., Miletic V., Rabson L.D., Smith J.A.,
RA Fearon D.T.;
RT *Identification of distinct C3b and C4b recognition sites in the
RT human C3b/C4b receptor (CRL, CD35) by deletion mutagenesis.";
RL J. Exp. Med. 166:1699-1717(1988).
RN [2]
RN SEQUENCE OF 503-2039 FROM N.A.
RX MEDLINE=87168191; PubMed=2951479;
RA Klickstein L.B., Wong W.W., Smith J.A., Weis J.H., Wilson J.G.,
RA Fearon D.T.;
RT "Human C3b/C4b receptor (CRL). Demonstration of long homologous
RT repeating domains that are composed of the short consensus repeats
RT characteristics of C3/C4 binding proteins.";
RL J. Exp. Med. 165:1095-1112(1987).
RN [3]
RN SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.
RX MEDLINE=86067975; PubMed=2933745;
RA Wong W.W., Klickstein L.B., Smith J.A., Weis J.H., Fearon D.T.;
RT *Identification of a partial cDNA clone for the human receptor for
RT complement fragments C3b/C4b.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:7111-7115(1985).
CC CC -1- FUNCTION: CRL, PRESENT ON ERYTHROCYTES, LEUKOCYTES, GLOMERULAR
CC PODOCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS, MEDIATES THE
CC BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE
CC ACTIVATED COMPLEMENT.
CC CC -1- SUBUNIT: MONOMER.
CC CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC CC -1- POLYMORPHISM: CRL IS RESPONSIBLE FOR THE KNOBS BLOOD GROUP SYSTEM.
CC CC -1- MISCELLANEOUS: SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A
CC LONG HOMOLOGOUS REPEAT (LHR). THE N-TERMINAL TWO SRCS OF LHR-A
CC CONTAINED A SITE DETERMINING C4 SPECIFICITY, AND THE N-TERMINAL
CC TWO SRCS OF LHR-B AND -C EACH HAD A SITE DETERMINING C3
CC SPECIFICITY.
CC CC -1- MISCELLANEOUS: THIS IS THE SEQUENCE OF THE F ALLOTYPE OF CRL.
CC CC -1- SIMILARITY: CONTAINS 30 SUSHI (SCR) DOMAINS.
CC CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; M11569; AAA52297.1; -
DR EMBL; M1617; AAA52298.1; -
DR EMBL; M1618; AAA52299.1; -
DR EMBL; Y00816; CAA68755.1; -
DR EMBL; X05309; CAA28933.1; -
DR PIR; A28507; A28507
DR PIR; A24748; A24748
DR PIR; B24748; B24748
DR PIR; C24748; C24748
DR PIR; S03843; S03843
DR HSP; P08603; 1HF1.
DR MIM; 120620; -
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 30.
DR SMART; SM00032; CGP; 30.
KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW Receptor; Sushi; Blood group antigen.
KW SIGNAL 1 41
FT CHAIN 42 2039 COMPLEMENT RECEPTOR TYPE 1.
FT DOMAIN 42 1971 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1972 1996 POTENTIAL.
FT DOMAIN 1997 2039 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 42 42 PYRROLIDONE CARBOXYLIC ACID (POTENTIAL).
FT DOMAIN 42 100 SUSHI A1.
FT DOMAIN 103 162 SUSHI A2.
FT DOMAIN 165 233 SUSHI A3.
FT DOMAIN 237 294 SUSHI A4.
FT DOMAIN 296 354 SUSHI A5.

```



```
FT DOMAIN 357 417 SUSHI A6.
FT DOMAIN 420 488 SUSHI A7.
FT DOMAIN 492 550 SUSHI B1.
FT DOMAIN 553 612 SUSHI B2.
FT DOMAIN 615 683 SUSHI B3.
FT DOMAIN 687 744 SUSHI B4.
FT DOMAIN 746 804 SUSHI B5.
FT DOMAIN 807 867 SUSHI B6.
FT DOMAIN 870 938 SUSHI B7.
FT DOMAIN 942 1000 SUSHI C1.
FT DOMAIN 1003 1062 SUSHI C2.
FT DOMAIN 1065 1133 SUSHI C3.
FT DOMAIN 1137 1194 SUSHI C4.
FT DOMAIN 1196 1254 SUSHI C5.
FT DOMAIN 1257 1317 SUSHI C6.
FT DOMAIN 1320 1388 SUSHI C7.
FT DOMAIN 1395 1453 SUSHI D1.
FT DOMAIN 1456 1515 SUSHI D2.
FT DOMAIN 1518 1586 SUSHI D3.
FT DOMAIN 1590 1647 SUSHI D4.
FT DOMAIN 1649 1707 SUSHI D5.
FT DOMAIN 1710 1770 SUSHI D6.
FT DOMAIN 1773 1841 SUSHI D7.
FT DOMAIN 1847 1905 SUSHI E1.
FT DOMAIN 1908 1966 SUSHI E2.
FT DISULFID 43 86 BY SIMILARITY.
FT DISULFID 73 99 BY SIMILARITY.
FT DISULFID 104 145 BY SIMILARITY.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 195 232 BY SIMILARITY.
FT DISULFID 238 280 BY SIMILARITY.
FT DISULFID 266 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 338 400 BY SIMILARITY.
FT DISULFID 386 416 BY SIMILARITY.
FT DISULFID 421 470 BY SIMILARITY.
FT DISULFID 450 487 BY SIMILARITY.
FT DISULFID 493 536 BY SIMILARITY.
FT DISULFID 523 549 BY SIMILARITY.
FT DISULFID 554 595 BY SIMILARITY.
FT DISULFID 581 611 BY SIMILARITY.
FT DISULFID 616 665 BY SIMILARITY.
FT DISULFID 645 682 BY SIMILARITY.
FT DISULFID 688 730 BY SIMILARITY.
FT DISULFID 716 743 BY SIMILARITY.
FT DISULFID 747 790 BY SIMILARITY.
FT DISULFID 776 803 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 866 BY SIMILARITY.
FT DISULFID 871 920 BY SIMILARITY.
FT DISULFID 900 937 BY SIMILARITY.
FT DISULFID 943 986 BY SIMILARITY.
FT DISULFID 973 999 BY SIMILARITY.
FT DISULFID 1004 1045 BY SIMILARITY.
FT DISULFID 1031 1061 BY SIMILARITY.
FT DISULFID 1066 1115 BY SIMILARITY.
FT DISULFID 1095 1132 BY SIMILARITY.
FT DISULFID 1138 1180 BY SIMILARITY.
FT DISULFID 1166 1193 BY SIMILARITY.
FT DISULFID 1197 1240 BY SIMILARITY.
FT DISULFID 1226 1253 BY SIMILARITY.
FT DISULFID 1258 1300 BY SIMILARITY.
FT DISULFID 1286 1316 BY SIMILARITY.
FT DISULFID 1321 1370 BY SIMILARITY.
FT DISULFID 1350 1387 BY SIMILARITY.
FT DISULFID 1396 1439 BY SIMILARITY.
FT DISULFID 1426 1458 BY SIMILARITY.
FT DISULFID 1457 1498 BY SIMILARITY.
FT DISULFID 1484 1514 BY SIMILARITY.
FT DISULFID 1519 1568 BY SIMILARITY.
FT DISULFID 1548 1585 BY SIMILARITY.
```

```
FT DISULFID 1591 1633 BY SIMILARITY.
FT DISULFID 1619 1646 BY SIMILARITY.
FT DISULFID 1650 1693 BY SIMILARITY.
FT DISULFID 1679 1706 BY SIMILARITY.
FT DISULFID 1711 1753 BY SIMILARITY.
FT DISULFID 1739 1769 BY SIMILARITY.
FT DISULFID 1774 1823 BY SIMILARITY.
FT DISULFID 1803 1840 BY SIMILARITY.
FT DISULFID 1848 1891 BY SIMILARITY.
FT DISULFID 1877 1904 BY SIMILARITY.
FT DISULFID 1909 1952 BY SIMILARITY.
FT DISULFID 1938 1965 BY SIMILARITY.
FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 509 509 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 860 860 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 959 959 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1028 1028 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1152 1152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1310 1310 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1481 1481 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1504 1504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1540 1540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1605 1605 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1763 1763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1908 1908 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2039 AA; 223589 MW; B2FD29B6AD3C5EB7 CRC64;
```

Query Match 21.9%; Score 255; DB 1: Length 2039;

Best Local Similarity 30.3%; Pred. No. 1,1,1,14;

Matches 67; Conservative 39; Mismatches 83; Indels 32; Gaps 13;

```
OY 3 CNE---LPPRRNTEILTGSMSDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGWALNPL 59
Db 43 CNAPELWLPARTNLT---DEFFPPIGYLNYECRPGYSGRPFSLCLKNSVWTGAKD- 97
OY 60 RKCQRPCGHPGDTPEGTFTLTGGNVFVGKAVYTCNKGYQLLGEINREC-----DTDG 115
Db 98 -RCRRKSRNPDPVNGMVHVING--IQFGSIKYSCTGYRLIGS-SSATCIISGDTVI 153
OY 116 WTNDIPICEVWC-LPVTAPENKIVSSAMEPDREYHFGQAVREVCNSG-----YKIEG 168
Db 154 WDNETPICDRIFCGLPPTI-TNGDFISTNRE---NFHYGSVVTYRCNPGSGGRKRVFELVG 209
OY 169 DEEMHC--SDD--GFWSKEKPKC-VEISCKSPDVINGSPIS 204
Db 210 EPSIVCTSNDDQVGWISGPAPOCIIPNKCTPPNVENGILVS 250
```

RESULT 6

```
LEM2_PIG ID LEM2_PIG STANDARD; PRT; 484 AA.
AC P98110;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE E-selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE (CD62E).
GN SELE.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
```

RC TISSUE-Aortic endothelium;
RX MEDLINE=95071392; PubMed=7526854;
RA Rollins S.A., Evans M.J., Johnson K.K., Elliot E.A., Squinto S.P.,
RA Matis L.A., Rother R.P.;
RT "Molecular and functional analysis of porcine E-selectin reveals a
RT potential role in xenograft rejection.";
RL Biochem. Biophys. Res. Commun. 204:763-771(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Aortic endothelium;
RX MEDLINE=94271236; PubMed=7516159;
RA Tsang Y.T.M., Haskard D.O., Robinson M.K.;
RT "Cloning and expression kinetics of porcine vascular cell adhesion
RT molecule.";
RL Biochem. Biophys. Res. Commun. 201:805-805(1994).
CC -!- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC ELAM-1 IS SIALYL-LEWIS X (ALPHA(1->3)FUCOSYLATED DERIVATIVES OF
CC POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC GLYCOPOLIPIDS).
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN ACUTE CELLULAR ALLOGRAFT
CC REJECTION AND PROBABLY ALSO IN XENOGRAFT REJECTION.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: TO OTHER SELECTINS/LECAMs.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS; PORCINE E-LECTIN LACKS
CC THE HUMAN SUSHI-1 AND -4 EQUIVALENTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L39076; AAA61545.1; -;
DR EMBL: U08350; AAA21541.1; -;
DR HSSP: P16581; LESL.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002396; Selectin.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00008; EGF; 1.
DR Pfam: PF00059; lectin_c; 1.
DR Pfam: PF00084; sushi; 4.
DR PRINTS: PR00343; SELECTIN.
DR SMART: SM00032; CCP; 4.
DR SMART: SM00034; CLECT; 1.
DR SMART: SM00181; EGF; 1.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE: PS00615; C-TYPE LECTIN_2; 1.
DR PROSITE: PS50041; C-TYPE LECTIN_2; 1.
KW Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW Selectin; Signal; Sushi; Repeat.
FT SIGNAL 1 22
FT CHAIN 23 484
FT DOMAIN 23 429
FT DOMAIN 430 451
FT DOMAIN 452 484
FT DOMAIN 39 139
FT DOMAIN 140 176
FT DOMAIN 180 236
FT DOMAIN 239 299
FT DOMAIN 302 362
FT DOMAIN 365 421
FT DISULFID 41 139
FT DISULFID 112 131
FT DISULFID 144 155
FT DISULFID 149 164
FT DISULFID 166 175

FT DISULFID 181 222
FT DISULFID 208 235
FT DISULFID 240 285
FT DISULFID 271 298
FT DISULFID 303 348
FT DISULFID 334 361
FT DISULFID 366 407
FT DISULFID 393 420
FT CARBOHYD 61 61
FT CARBOHYD 65 65
FT CARBOHYD 79 79
FT CARBOHYD 160 160
FT CARBOHYD 201 201
FT CARBOHYD 254 254
FT CARBOHYD 376 376
FT CARBOHYD 400 400
FT CONFLICT 253 253
FT CONFLICT 313 313
FT CONFLICT 321 321
FT CONFLICT 327 327
FT CONFLICT 363 363
FT CONFLICT 384 384
FT CONFLICT 461 484
SQ SEQUENCE 484 AA; 52567 MW; AFF74FE25C1FD013 CRC64;
Query Match 21.8%; Score 253.5; DB 1; Length 484;
Best Local Similarity 28.6%; Pred. No. 3;le-15;
Matches 56; Conservative 36; Mismatches 83; Indels 21; Gaps 9;
QY 22 QTYPEGTQAIYKCRPGYRSLGNIMVC-RKGEWVALNLRKCKRKC---GHPCDTPFGT 77
DB 196 QSLPWNTTCAFEKGEFELIGPHLQCTSSGSGWDGKKP--TCRAVTCDTVGHFQN---GD 250
QY 78 FTLTGNVFYGVKAV--YTCNEGYQLLGEINRECDTDG-WTNDIPICEVWKLVPVAP 134
DB 251 VSCNHSISGEFAVYKSTCHFTCAEFGLOGPAQI-ECTAQCQWQTQAPVCAVCAVSP 309
QY 135 ENGIYSSAMEPDREYHFGQAVRVCNCGYKIEGDEMHCSDDGFWNSKEPKVCVEISCKS 194
DB 310 KNG-LVKFTHSPGTGEFTYKSSCAFSCEEGFELRGSQAQLACTSQGQMTQVEYPCVQVCCS 368
QY 195 PDV-----INGSPI 203
DB 369 LEVPREINMSSCGEPV 384
RESULT 7
ID DAF_PONPY STANDARD; PRT; 340 AA.
AC P49457;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Complement decay-accelerating factor (CD55) (Fragment).
GN DAF OR CD55.
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94110622; PubMed=7506731;
RA Nickells M.W., Alvarez J.I., Lublin D.M., Atkinson J.P.;
RT "Characterization of DAF-2, a high molecular weight form of decay-
RT accelerating factor (DAF; CD55), as a covalently cross-linked dimer
RT of DAF-1.";
RL J. Immunol. 152:676-685(1994).
CC -!- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT
CC CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT
CC C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.
CC INTERACTION OF DAF WITH CELL-ASSOCIATED C4B AND C3B POLYPEPTIDES

DR EMBL; D55659; BAA09516.1; JOINED.
DR EMBL; D55660; BAA09516.1; JOINED.
DR EMBL; D55661; BAA09516.1; JOINED.
DR EMBL; D55662; BAA09516.1; JOINED.
DR EMBL; D55663; BAA09516.1; JOINED.
DR EMBL; D55664; BAA09516.1; JOINED.
DR EMBL; D55665; BAA09516.1; JOINED.
DR EMBL; D55667; BAA09517.1; -.
DR EMBL; D55656; BAA09517.1; JOINED.
DR EMBL; D55657; BAA09517.1; JOINED.
DR EMBL; D55658; BAA09517.1; JOINED.
DR EMBL; D55659; BAA09517.1; JOINED.
DR EMBL; D55660; BAA09517.1; JOINED.
DR EMBL; D55661; BAA09517.1; JOINED.
DR EMBL; D55662; BAA09517.1; JOINED.
DR EMBL; D55663; BAA09517.1; JOINED.
DR EMBL; D55664; BAA09517.1; JOINED.
DR EMBL; D55665; BAA09517.1; JOINED.
DR EMBL; D55666; BAA09518.1; JOINED.
DR EMBL; D55667; BAA09518.1; -.
DR EMBL; D55656; BAA09518.1; JOINED.
DR EMBL; D55657; BAA09518.1; JOINED.
DR EMBL; D55658; BAA09518.1; JOINED.
DR EMBL; D55659; BAA09518.1; JOINED.
DR EMBL; D55660; BAA09518.1; JOINED.
DR EMBL; D55661; BAA09518.1; JOINED.
DR EMBL; D55662; BAA09518.1; JOINED.
DR EMBL; D55663; BAA09518.1; JOINED.
DR EMBL; D55664; BAA09518.1; JOINED.
DR EMBL; D55665; BAA09518.1; JOINED.
DR EMBL; D55666; BAA09519.1; -.
DR EMBL; D55667; BAA09519.1; -.
DR EMBL; D55656; BAA09519.1; JOINED.
DR EMBL; D55657; BAA09519.1; JOINED.
DR EMBL; D55658; BAA09519.1; JOINED.
DR EMBL; D55659; BAA09519.1; JOINED.
DR EMBL; D55660; BAA09519.1; JOINED.
DR EMBL; D55661; BAA09519.1; JOINED.
DR EMBL; D55662; BAA09519.1; JOINED.
DR EMBL; D55663; BAA09519.1; JOINED.
DR EMBL; D55664; BAA09519.1; JOINED.
DR EMBL; D49416; BAA08396.1; -.
DR EMBL; D49417; BAA08397.1; -.
DR EMBL; D49418; BAA08398.1; -.
DR EMBL; D49419; BAA08399.1; -.
DR EMBL; D49420; BAA08400.1; -.
DR EMBL; D49421; BAA08401.1; -.
DR HSP; P08603; LHCC
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
KW Complement pathway; Glycoprotein; Membrane; Repeat; GPI-anchor;
KW Alternative splicing; Signal; Sushi.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 451 COMPLEMENT DECA-ACCELERATING FACTOR.
FT PROPEP 452 507 REMOVED IN MATURE FORM (BY SIMILARITY).
FT DOMAIN 32 93 SUSHI 1.
FT DOMAIN 94 158 SUSHI 2.
FT DOMAIN 159 220 SUSHI 3.
FT DOMAIN 221 283 SUSHI 4.
FT DOMAIN 284 458 SER/THR-RICH.
FT LIPID 451 451 GPI-ANCHOR (BY SIMILARITY).
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 442 442 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 302 334 MISSING (IN ISOFORM GDA-TCS).
FT VARSPLIC 335 366 MISSING (IN ISOFORM GDA-TCS).
FT FT MISSING (IN ISOFORM GDA-TCS, ISOFORM GDA-
FT FT TCS AND ISOFORM GDA-SEC).
FT FT MISSING (IN ISOFORM GDA-TCS AND GDA-
FT FT SEC).

FT VARSPLIC 425 458 MISSING (IN ISOFORM GDA-B-GPI, ISOFORM
FT FT GDA-B-TCL, ISOFORM GDA-B-TCL AND ISOFORM
FT FT GDA-B-SEC).
FT FT THYKVDTSFACGASHNLADIADAKEDLRDRFSNAQNISSLLQ
FT FT VLGAQAQTQ -> GHMCILKTLVLLVVLVIIG (IN
FT FT ISOFORM GDA-B-GPI).
FT FT THYKVDTSFACGASHNLADIADAKEDLRDRFSNAQNISSLLQ
FT FT VLGAQAQTQ -> DTCV (IN ISOFORM GDA-B-SEC).
FT FT THYKVDTSFACGASHNLADIADAKEDLRDRFSNAQNISSLLQ
FT FT VLGAQAQTQ -> ANRHRTKNPLIYVIT (IN ISOFORM
FT FT GDA-TCS AND ISOFORM GDA-B-TCS).
SQ SEQUENCE 507 AA; 55263 MW; D25BBB7749425210 CRC64;

Query Match 21.2%; Score 246.5; DB 1; Length 507;
Best Local Similarity 30.4%; Pred. No. 1.4e-14;
Matches 66; Conservative 33; Mismatches 83; Indels 35; Gaps 11;

QY 1 EDCN---ELPPRRNTEILTGSMSDQTY-PEGTOAIYKCRPGYRS---LGNVIMVCRKG-E 52
DB 90 EFCNRSCEVPRLPYAILKSAISKYFPDVTVEYECRPGYRKYPYQSGKITCRDDL 149
QY 53 WVALNPLRCKKRCPCGHPGDPFGFTLT-----GNVFYGVKAVYTCNEGYQLLEIN 107
DB 150 W--STPAEFCEKQCPNPGELVNGINVTDLGLSQIF-----FSCDPCGYRLTGEAS 200
QY 108 ---YRECDTGTNDIPICEVVKCLPVTAPENGKLTSSAMEPDREYHFCQAVRFVCSNY 164
DB 201 AFCMTKGNAGVSSSLPTCIKICPEPOIENGRIVNE-----EDTYEYRHVVYTCNKG 256
QY 165 KIEGDEEMHC---SDGFWSEKPKCKVEISCKSPDVI 198
DB 257 VLTGKSHISCIVRDDVDGEMSDPPPTC---RVKSPPIV 290

RESULT 9
VCP_VACCV STANDARD; PRT; 263 AA.
AC P10998;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement control protein precursor (VCP) (Secretory protein 35)
DE (Protein C3) (28 kDa protein).
GN C3L
OS Vaccinia virus (strain WR), and
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10254, 10249;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-37.
RC STRAIN-WR;
RX MEDLINE=88318974; PubMed=3412473;
RA Kotwal G.J., Moss B.;
RT "Vaccinia virus encodes a secretory polypeptide structurally related
RL to complement control proteins.";
RN [2]
RP Nature 335:176-178(1988).
RP SEQUENCE FROM N.A.
RC STRAIN-WR;
RX MEDLINE=89073756; PubMed=2849238;
RA Kotwal G.J., Moss B.;
RT "Analysis of a large cluster of nonessential genes deleted from a
RL vaccinia virus terminal transposition mutant.";
RN [3]
RP Virology 167:524-537(1988).
RP SEQUENCE FROM N.A.
RC STRAIN-COPENHAGEN;
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RT "The complete DNA sequence of vaccinia virus.";

RC TISSUE-Testis;
RX MEDLINE-931119658; PubMed-8418811;
RA Cervoni F., Fenichel P., Ahoundi C., Hsi B.L., Rossi B.;
RT "Characterization of a cDNA clone coding for human testis membrane
RT cofactor protein (MCP, CD46).";
RL Mol. Reprod. Dev. 34:1107-1113(1993).
[3]
RN SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE-91267562; PubMed-2050389;
RA Purcell D.F., Russell S.M., Deacon N.J., Brown M.A., Hooker D.J.,
RA McKenzie I.F.;
RT "Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a
RT regulator of complement activation.";
RL Immunogenetics 33:335-344(1991).
[4]
RN SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE-94014356; PubMed-7691939;
RA Cui W., Hourcade D., Post T., Greenlund A.C., Atkinson J.P.,
RA Kumar V.;
RT "Characterization of the promoter region of the membrane cofactor
RT protein (CD46) gene of the human complement system and comparison to
RT a membrane cofactor protein-like genetic element.";
RL J. Immunol. 151:4137-4146(1993).
[5]
RN ALTERNATIVE SPLICING.
RX MEDLINE-92289809; PubMed-1601037;
RA Russell S.M., Sparrow R.L., McKenzie I.F.C., Purcell D.F.J.;
RT "Tissue-specific and allelic expression of the complement regulator
RT CD46 is controlled by alternative splicing.";
RL Eur. J. Immunol. 22:1513-1518(1992).
CC -!- FUNCTION: MCP IS AN IMPORTANT MEMBRANE PROTEIN FOR PROTECTING HOST
CC CELLS FROM DAMAGE BY COMPLEMENT. MCP POSSESSES COFACTOR ACTIVITY
CC FOR INACTIVATION OF C3B AND C4B BY SERUM FACTOR I. THIS WOULD
CC ALLOW MCP TO REGULATE THE COMPLEMENT SYSTEM AT THE LEVEL OF THE C3
CC CONVERTASES. MAY BE INVOLVED IN THE FUSION OF THE SPERMATOZOON WITH
CC THE OOCYTE DURING FERTILIZATION. MAY ACT BY PREVENTING COMPLEMENT
CC ACTIVATION AT THE FETO-MATERNAL INTERFACE ON THE
CC SYNCYTOTROPHOBLAST LAYER OF PLACENTA.
CC -!- SURCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 14 FORMS ARE PRODUCED BY
CC ALTERNATIVE SPLICING. THE FORM SHOWN IS THE TRANSCRIPT A. THE
CC GAMMA ISOFORM (TRANSCRIPTS A AND B) IS PREFERENTIALLY EXPRESSED IN
CC EBV-B CELLS AND LEUKEMIC CELLS. THE 66 kDa ALPHA ISOFORM AND THE
CC 56 kDa BETA ISOFORM FOUND IN ALL TISSUES EXCEPT SPERM CORRESPOND
CC TO THE TRANSCRIPTS C AND D, AND THE TRANSCRIPTS E AND F
CC RESPECTIVELY. THE EXON 9 IS SPECIFICALLY DELETED IN SOME
CC PLACENTAE. THE EPSILON ISOFORM CORRESPONDS TO THE TRANSCRIPTS I
CC AND J. SPERMATOZOON DELTA ISOFORM (TRANSCRIPT N) DELETE BOTH EXONS
CC 12 AND 13. ALL TISSUES DIFFERENTIALLY SPLICED EXON 13.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY ALMOST ALL CELLS EXCEPT
CC ERYTHROCYTES AND SOME BONE MARROW CELLS.
CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED IN THE SER/THR-RICH DOMAIN
CC (PROBABLY).
CC -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD46 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd46.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y00651; CAA68675.1; -
CC EMBL; S51940; AAB24802.1; -
CC EMBL; M58050; AAG62833.1; -
CC EMBL; A18585; CAA01400.1; -
CC EMBL; S65879; AAD13968.1; -
CC PIR; S01896; S01896.
CC HSP; P10996; 1VVD.
CC MIN; P10920; -

DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 4.
DR SMART: SM00032; CCP; 4.
KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW Sushi; Alternative splicing.
FT SIGNAL 1 34
FT CHAIN 35 377 MEMBRANE COFACTOR PROTEIN.
FT DOMAIN 35 328 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 329 351 POTENTIAL.
FT DOMAIN 352 377 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 95 SUSHI 1.
FT DOMAIN 98 158 SUSHI 2.
FT DOMAIN 161 224 SUSHI 3.
FT DOMAIN 227 284 SUSHI 4.
FT DOMAIN 287 311 SER/THR-RICH.
FT DISULFID 35 80 BY SIMILARITY.
FT DISULFID 64 94 BY SIMILARITY.
FT DISULFID 99 141 BY SIMILARITY.
FT DISULFID 127 157 BY SIMILARITY.
FT DISULFID 162 210 BY SIMILARITY.
FT DISULFID 191 223 BY SIMILARITY.
FT DISULFID 228 270 BY SIMILARITY.
FT DISULFID 256 283 BY SIMILARITY.
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 286 300 MISSING (IN A SECOND FORM).
FT VARSPLIC 340 352 MISSING (IN ISOFORM M).
FT VARSPLIC 353 361 YLQRRKKG -> DIFKGGRRKGGKQMVNLNPLTRLNQPLQ
FT VARSPLIC 362 377 TLTDETHREVFETSL -> KADGGAEYATYQKSTTPAEQ
FT VARSPLIC 362 377 RG (IN ISOFORM B/D/F/H/J/L).
FT VARSPLIC 362 377 MISSING (IN ISOFORM M AND ISOFORM N).
FT VARSPLIC 340 361 MISSING (IN ISOFORM M).
FT VARSPLIC 339 339 I -> IGKQMVNLNPLTRLNQPLQDSREAE (IN
FT ISOFORM N).
SQ SEQUENCE 377 AA; 42247 MW; 2CA6F61752570B57 CRC64;
Query Match 20.5%; Score 238.5; DB 1; Length 377;
Best Local Similarity 29.6%; Pred. No. 5e-14;
Matches 64; Conservative 32; Mismatches 99; Indels 21; Gaps 9;
QY 2 DCNELPPRNTTEILTGSWSDQTYPECTQAIYKCRPY----RSLGNVIMVCRKGEWALNP 58
DB 33 DACEEPTFEAMELIGK-PKPYEIGERYDYKCKGKYFIPPLATHITCDRNHTLTPVSD 91
QY 59 LRKCKRCPGHPGDTFGFTLTGGNVVEYGVKAVYTCNEGYYOLLG-EINYRECDTDG-- 115
DB 92 -DACYRETCPYIRD-PLNGQAVPANGTYEFGYQMHFICNEGYYLIGEEIY--CEKGSV 147
QY 116 --WTNDIPICEVYVKLPVTPAPENGKIVVSAMEPDRYHFGQAVRFVCNSG-----YKIEG 168
DB 148 AIWSGKPPICEKVLCTPPPKIKNGKHTFEVE--VFEYLDVATYSCDPAGPDPFSLIG 204
QY 169 DEEMHCSDGCFNSKEPKCKVEISCKSPDVINGSPIS 204
DB 205 ESTIYCGDNVMSRAAPECKVCKRFPVVGKQIS 240
RESULT 12
CCPH_HSVSA
ID CCPH_HSVSA STANDARD; PRT; 360 AA.
AC Q01016;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Complement control protein homolog precursor (CCPH).
GN 4 OR CCPH.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OC NCBI_TaxID-10383;


```

[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE-92333688; PubMed-1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058(1992).
[2]
RN RP SIMILARITY TO CCP.
RX MEDLINE-92260674; PubMed-1316492;
RA Albrecht J.-C., Fleckenstein B.;
RT "New member of the multigene family of complement control proteins in
herpesvirus saimiri.";
RL J. Virol. 66:3937-3940(1992).
CC -|- ALTERNATIVE PRODUCTS: A MEMBRANE-BOUND FORM AND A SECRETED FORM
CC ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC -|- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF
CC COMPLEMENT ACTIVATION (RCA).
CC -|- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X64346; CAA45626.1; -
DR EMBL; X64346; CAA45627.1; -
DR EMBL; X60283; CAA42823.1; -
DR EMBL; X60283; CAA42822.1; -
DR PIR; B42534; WMBE2E.
DR PIR; A42534; WMBE1E.
DR PIR; S24567; S24567.
DR HSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
KW Signal; Repeat; Sushi; Transmembrane; Alternative splicing;
KW Glycoprotein.
FT SIGNAL 1 20
FT CHAIN 21 360
FT DOMAIN 83 143
FT DOMAIN 146 206
FT DOMAIN 209 265
FT TRANSMEM 328 350
FT DISULFID 84 125
FT DISULFID 111 142
FT DISULFID 147 191
FT DISULFID 175 205
FT DISULFID 210 252
FT DISULFID 238 264
FT CARBOHYD 36 36
FT CARBOHYD 39 39
FT CARBOHYD 46 46
FT CARBOHYD 72 72
FT CARBOHYD 155 155
FT CARBOHYD 294 294
FT VARSPLIC 289 302
FT VARSPLIC 303 360
FT VARSPLIC 360 AA; 40006 MW; 6278A6C2ECD49669 CRC64;
SQ SEQUENCE 360 AA; 40006 MW; 6278A6C2ECD49669 CRC64;

Query Match 20.3%; Score 236; DB 1; Length 360;
Best Local Similarity 33.3%; Pred. No. 8e-14;
Matches 61; Conservative 17; Mismatches 89; Indels 16; Gaps 7;
OY 24 YPEGQAIYKCRPGYRSLGNVIMVCRKGEWALNPLRKQKRCRPGDPTPGTFTLGG 83
DB 44 Y:NGTTLHTWCREGYAKRPVQITVCVNGWTV---PKCKQKKKCTPQDLLNGRTVT-G 99
```

```

OY 84 NVPEYGVKAVYTCNEGYYQLIGEIN----YRECDTDGWTNDIPICEVVKVCLPVPAPENGKI 139
DB 100 LLY-YGSVITYTCNSCYSLIGSTTSACLLKRGGRVDMTPRPIDIKKCKPPQIANG-- 156
OY 140 VVSAMPDREYHFGQAVRFVCSNGYK--TEGDEEMHCSDGFW-SKEKPKCVISCKSPD 196
DB 157 --THTNVKDFYTYLDTVTYSCNDETKLTLTGPSKCLCSETGSWVPNGETKCEFIKCLKPQ 214
OY 197 VIN 199
DB 215 VAN 217

RESULT 13
DAFL_MOUSE
ID DAFL_MOUSE STANDARD; PRT; 390 AA.
AC Q61475; Q61397; P97732;
CT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Complement decay-accelerating factor, GPI-anchored precursor
DE (DAF-GPI).
GN DAF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-Testis;
RX MEDLINE-95403982; PubMed-7545711;
RA Spicer A.P., Seldin M.F., Gendler S.J.;
RT "Molecular cloning and chromosomal localization of the mouse decay-
RT accelerating factor genes. Duplicated genes encode
RT glycosylphosphatidylinositol-anchored and transmembrane forms.";
RL J. Immunol. 155:3079-3091(1995).
RN [2]
RP SEQUENCE OF 7-390 FROM N.A.
RC STRAIN-BALB/C; TISSUE-Spleen;
RX MEDLINE-96362213; PubMed-8671624;
RA Fukuoaka Y., Yasui A., Okada N., Okada H.;
RT "Molecular cloning of murine decay accelerating factor by
RT immunoscreening.";
RL Int. Immunol. 8:379-385(1996).
CC -|- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY
CC SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -|- TISSUE SPECIFICITY: BRAIN, SECRETORY EPITHELIA, SKELETAL MUSCLE,
CC LIVER, TESTES, THYMUS, SPLEEN AND LYMPH NODE.
CC -|- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
CC ACTIVE SITE ON SCR3 (BY SIMILARITY).
CC -|- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC -|- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC (RCA) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; L41366; AAB00091.1; -
DR EMBL; D63679; BAA09830.1; -
DR HSSP; P08603; IHCC.
DR MGD; MGI:104850; Dafi.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
KW Complement pathway; Glycoprotein; Membrane; Repeat; GPI-anchor;
```


WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd62e.htm".

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M30640; AAA52377.1; -
 EMBL; M61893; AAA52375.1; -
 EMBL; M61895; AAA52375.1; JOINED.
 EMBL; M61887; AAA52375.1; JOINED.
 EMBL; M61888; AAA52375.1; JOINED.
 EMBL; M61890; AAA52375.1; JOINED.
 EMBL; M61891; AAA52375.1; JOINED.
 EMBL; M61892; AAA52375.1; JOINED.
 EMBL; M4736; AAA52376.1; -
 PIR; A32606; A32606.
 PIR; A35046; A35046.
 PIR; A38615; A38615.
 PDB; 1ESL; 31-AUG-94.
 PDB; 1KJA; 03-APR-96.
 MIM; 131210; -
 InterPro: IPR000561; EGF-like.
 InterPro: IPR002396; Selection.
 InterPro: IPR000436; Sushi_SCR_CCP.
 InterPro: IPR001304; lectin_c.
 Pfam; PF00008; EGF_1.
 Pfam; PF00059; lectin_c; 1.
 Pfam; PF00084; sushi; 6.
 PRINTS; PR00343; SELECTION.
 SMART; SM00032; CCP; 6.
 SMART; SM00034; CLECT; 1.
 SMART; SM00181; EGF; 1.
 PROSITE; PS00022; EGF_1; 1.
 PROSITE; PS01186; EGF_2; 1.
 PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
 PROSITE; PS0041; C-TYPE_LECTIN_2; 1.
 Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
 KW Selection; Signal; Sushi; Repeat; Polymorphism; 3D-structure.
 FT SIGNAL 1 21
 FT CHAIN 22 610 E-SELECTIN.
 FT DOMAIN 22 556 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 557 578 POTENTIAL.
 FT DOMAIN 579 610 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 139 175 C-TYPE LECTIN (SHORT FORM).
 FT DOMAIN 179 238 EGF-LIKE.
 FT DOMAIN 241 300 SUSHI 1.
 FT DOMAIN 303 363 SUSHI 2.
 FT DOMAIN 366 426 SUSHI 3.
 FT DOMAIN 429 489 SUSHI 4.
 FT DOMAIN 492 548 SUSHI 5.
 FT DISULFID 40 138 SUSHI 6.
 FT DISULFID 111 130
 FT DISULFID 143 154
 FT DISULFID 148 163
 FT DISULFID 165 174
 FT DISULFID 180 224
 FT DISULFID 210 237
 FT DISULFID 242 286
 FT DISULFID 272 299
 FT DISULFID 304 349
 FT DISULFID 335 362
 FT DISULFID 367 412
 FT DISULFID 398 425
 FT DISULFID 430 475
 FT DISULFID 461 488
 FT DISULFID 493 534
 FT DISULFID 520 547
 FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 160 160 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 203 203 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 265 265 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 312 312 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 527 527 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 130 130 C -> W (IN DBSNP:5360).
 FT VARIANT 149 149 /FTID-VAR_011790.
 FT VARIANT 149 149 S -> R (ASSOCIATED WITH A RISK FACTOR FOR CAD).
 FT VARIANT 295 295 /FTID-VAR_004191.
 FT VARIANT 421 421 E -> K (IN DBSNP:5364).
 FT VARIANT 421 421 /FTID-VAR_011791.
 FT VARIANT 468 468 E -> Q (IN DBSNP:5366).
 FT VARIANT 468 468 /FTID-VAR_011792.
 FT VARIANT 575 575 H -> Y (IN DBSNP:5368).
 FT VARIANT 575 575 /FTID-VAR_011793.
 FT VARIANT 575 575 L -> F (IN DBSNP:5355).
 FT VARIANT 575 575 /FTID-VAR_011794.
 FT SEQUENCE 610 AA; 66655 MM; 7D43E3C0D1229229 CRC64;
 SQ
 Query Match 19.8%; Score 230.5; DB 1; Length 610;
 Best Local Similarity 27.5%; Pred. No. 4.3e-13;
 Matches 53; Conservative 30; Mismatches 81; Indels 29; Gaps 10;
 QY 23 TYPEGTQAIYKCRPGYRSLGNVIMVC-RKGEWVALNPLRK---CQ---KRP-----CGHP 70
 DB 261 SFPWNITCTDCERGFELMGAQSLQCTSSGNWMDNEKPTCAVTCRAVYRQPGQSVRCSH- 319
 QY 71 GDTPEFTTTLTGNVFEYGVKAVYTCNEGVQLLGEINRYECDDTG-WTNDIPICEVVKCL 129
 DB 320 --SPAGEFT-----FKSSCNFTCEGFMLOGPAQV-ECTTGQWTOQIPVCEAFQCT 368
 QY 130 PVTAPENGKIVSSAMEPDREYHFGQAVRVCNSGYKTEGDEMHCSDDGFWSKKPKVCE 189
 DB 369 ALSNPERG-YMNCPLPSAGSFRYSCSFCEQGFVLKSGKRLOCGTGEWNEKPTCEA 427
 QY 190 ISCKSPDVINGSP 202
 DB 428 VRC---DAVHOPP 437
 RESULT 15
 ID C4BP_MOUSE STANDARD; PRT; 469 AA.
 AC P08607;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-OCT-2001 (Rel. 40, Last annotation update)
 DE C4b-binding protein precursor (C4bp).
 GN C4BPA OR C4BP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88024997; PubMed=3663616;
 RA Kristensen T., Ogata R.T., Chung L.P., Reid K.B.M., Tack B.F.;
 RT "cDNA structure of murine C4b-binding protein, a regulatory component
 of the serum complement system.";
 RL Biochemistry 26:4668-4674(1987).
 CC -I- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
 CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
 CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
 CC ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3
 CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
 CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:01:39 ; Search time 67.22 Seconds
(without alignments)
532.728 Million cell updates/sec

Title: US-09-316-163-9
Perfect score: 1163
Sequence: 1 EDCNELPPRRNTEILTGSWS.....VEISCKSPDVINGSPISOKI 207

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1163	100.0	1172	4 Q9NU87	Q9nu87 homo sapien
2	848	72.9	1236	11 Q91YB6	Q91yb6 rattus norv
3	499	42.9	669	6 Q28085	Q28085 bos taurus
4	368.5	31.7	1053	13 Q91275	Q91275 paralabrax
5	276.5	23.8	645	12 Q9WRU2	Q9wru2 macaca mula
6	275	23.6	360	12 Q9YRQ8	Q9yrt8 ateline her
7	270.5	23.3	550	12 P88903	P88903 kaposi's sa
8	270.5	23.3	550	12 O40912	O40912 kaposi's sa
9	269.5	23.2	395	12 Q9J2M6	Q9j2m6 macaca mula
10	263.5	22.7	3389	4 Q96QU9	Q96qu9 homo sapien
11	263.5	22.7	3508	4 Q96RM4	Q96rm4 homo sapien
12	261.5	22.5	533	4 Q9H4W4	Q9h4w4 homo sapien
13	260	22.4	679	11 Q99254	Q99254 mus musculus
14	260	22.4	3567	11 Q9ES77	Q9es77 mus musculus
15	258.5	22.2	3564	11 Q923L3	Q923l3 mus musculus
16	257	22.1	347	6 Q9MYJ6	Q9myj6 pan troglod

17	256	22.0	522	6 Q28769	Q28769 papio cynoc
18	255.5	22.0	259	12 P87616	P87616 cowpox viru
19	255	21.9	559	4 Q9UQV2	Q9uqv2 homo sapien
20	255	21.9	2039	4 Q16745	Q16745 homo sapien
21	255	21.9	2489	4 Q16744	Q16744 homo sapien
22	253.5	21.8	482	6 Q28982	Q28982 sus scrofa
23	249	21.4	347	6 Q9MYJ7	Q9myj7 gorilla gor
24	248	21.3	315	6 Q28770	Q28770 papio cynoc
25	248	21.3	343	6 Q9MYJ4	Q9myj4 papio hamad
26	248	21.3	363	6 O02839	O02839 sus scrofa
27	247	21.2	661	6 Q29531	Q29531 pan troglod
28	247	21.2	2014	6 Q29530	Q29530 pan troglod
29	246.5	21.2	305	6 Q9MYJ5	Q9myj5 pan troglod
30	244	21.0	343	6 Q9MYJ2	Q9myj2 erythrocebu
31	244	21.0	343	6 Q9MYJ9	Q9myj9 erythrocebu
32	243.5	20.9	497	11 Q63612	Q63612 rattus norv
33	243.5	20.9	559	11 Q63135	Q63135 rattus norv
34	243	20.9	222	6 O19122	O19122 callimico q
35	243	20.9	560	5 Q22328	Q22328 caenorhabdi
36	241	20.7	343	6 Q9MYJ3	Q9myj3 macaca mula
37	240.5	20.7	271	6 Q9MYJ0	Q9myj0 erythrocebu
38	240.5	20.7	483	11 Q64735	Q64735 mus musculu
39	239.5	20.6	481	4 Q9H284	Q9h284 homo sapien
40	239.5	20.6	1911	6 Q29528	Q29528 papio hamad
41	238.5	20.5	216	12 Q98VL5	Q98vl5 monkeypox v
42	238.5	20.5	349	4 Q15429	Q15429 homo sapien
43	238.5	20.5	378	6 O62837	O62837 saguinus oe
44	237	20.4	222	6 O19128	O19128 pithecia pi
45	237	20.4	1653	5 Q9VIU9	Q9viu9 drosophila

ALIGNMENTS

RESULT 1

Q9NU87 PRELIMINARY; PRT: 1172 AA.

AC Q9NU87;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE DJ177P10.1.1 (H FACTOR 1 (COMPLEMENT) ISOFORM 1).
 GN HFL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bird C.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL049744; CAB70597.1;
 DR HSSP; P08603; 1HFH.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00084; Sushi; 19.
 DR SMART; SM00032; CCP; 19.
 SQ SEQUENCE 1172 AA; 132087 MW; 8F5B954C4B4FA454 CRC64;

Query Match 100.0%; Score 1163; DB 4; Length 1172;
 Best Local Similarity 100.0%; Pred. No. 3.9e-110;
 Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDCNELPPRRNTEILTGSWSDDTYPEGTQAIYKCPGYBSLGNVIMVCRGEMVNLPLR 60
 Db 19 EDCNELPPRRNTEILTGSWSDDTYPEGTQAIYKCPGYBSLGNVIMVCRGEMVNLPLR 78
 QY 61 KCQKRPCGHPGDPFGTETLTGNNVFYGVKAVYTCNEGYYQLLGEINYTECDTGTNDI 120
 Db 79 KCQKRPCGHPGDPFGTETLTGNNVFYGVKAVYTCNEGYYQLLGEINYTECDTGTNDI 138
 QY 121 PICEVVKLPVTAPNGKIVSSAMEPDREYHFGQAVRFVCSNGYKIEGDEMHCSDDGF 180

```

Db      139  PICEVYVKCLPVTAPENCKIVSSNAMEDPREYHFCQAVRFVNCVSGYKIEGDEMHKSDDGFW 198
Qy      181  SKEPKPCVEISCKSPDVINGSPISOKI 207
      |||||
Db      199  SKEPKPCVEISCKSPDVINGSPISOKI 225
      |||||

RESULT  2
Q11YB6  Q91YB6  PRELIMINARY;      PRT; 1236 AA.
AC      Q91YB6;
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE      DECOMPLEMENT INHIBITORY FACTOR H.
GN      FH.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RS      SEQUENCE FROM N.A.
RC      STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RA      Demberg T., Goetze O., Schlaf G.;
RT      "Rat complement factor H: molecular cloning, sequencing and expression
RT      in tissues and isolated cells.";
RL      Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AJ320522; CAC67513.1;
SQ      SEQUENCE 1236 AA; 140343 MW; 1AC89FFA28232EBF CRC64;

```

Query Match	72.9%	Score 348;	DB 11;	Length 1236;
Best Local Similarity	71.4%;	Pred. No. 6.9e-78;		
Matches 147;	Conservative 21;	Mismatches 38;	Indels 0;	Gaps 0;
Qy	1	EDCNELPRRTEILTGTSGSDQTPEGTQATYKCRPGYRSILGNVIMCRKGEWALNPLR	60	
Db	19	EDCGPPPRENSEILSGSMSEQLYSEGTQATYKCRPGYRTLTIVKVCCKNGEWPNSFR	78	
Qy	61	KQKRPCGHPGDTTFFGTTLTGNVFEYGVKAVYTCNCGYOLLGGINTECDTGWNTDI	120	
Db	79	ICRKRPCGHPGDTTFFGSRFLAVGSEFEAGVYVTCDECYOLLGIDYRECDADGWNTDI	138	
Qy	121	PICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRVCNMSGYKIEGDEEMHCSDGCGW	180	
Db	139	PICEVVKCLPTELENGRIVSGAAEPDQEYFYGQVVRFCNMSGFKIEGCKEMHCSENGLW	198	
Qy	181	SKRKPCKVEISCKSPDIVINGSPISQK	206	
Db	199	SNEKPQCVIEISCLPPRVENGDIYLVK	224	

```

RESULT      3
Q28085
ID          Q28085      PRELIMINARY;      PRT;      669 AA.
AC          Q28085;
DT          01-NOV-1996 (TrEMBLrel. 01, Created)
DT          01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT          01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE          CCP MODULES 3-12, WITH PARTS OF CCP 2 AND 13 (FRAGMENT).
OS          Bos taurus (Bovine).
OC          Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC          Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae.
OC          Bovidae; Bovinae; Bos.
OX          NCBI_TaxID=9913;
RN          [1]
RC          SEQUENCE FROM N.A.
RP          TISSUE=LIVER;
RT          MEDLINE=96202005; PubMed=8615824;
RA          Soames C.J., Day A.J., Sim R.B.;
RL          "Prediction from sequence comparisons of residues of factor H involved
RL          in the interaction with complement component C3b.";
RL          Biochem. J. 315:523-531(1996).

```

```

DR EMBL; X98697; CAA67257.1; .
DR HSSP; P10998; lVVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 11.
DR SMART; SM00032; CCP; 11.
FT NON_TER      1
FT NON_TER      669
SQ SEQUENCE      669 AA; 75683 MW; D0D9DB30EE747AC2 CRC64;

Query Match          42.9%; Score 499; DB 6; Length 669;
Best Local Similarity 64.9%; Pred. No. 1.6e-42;
Matches 85; Conservative 19; Mismatches 27; Indels 0; Gaps 0

Qy    76   GTFTLTGNVFEYGVKAVYTTCNCGYQLLGELTNYRECDDTGWTNDIPICEVVKKLPVTPE 135
       1 : - - - - - | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     3   GSPIHLAEGNFQEYGAKVYYTCDEGTQWGMENFRECDDTGWTNDIPICEVVKKLPVTPE 62
       1 : - - - - - | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy    136  NGKITVSAMEPDREYHFHQAVRFVCNCSGYKIEGDSEMHCSDDGFWSKPKCYEISCKSP 195
       1 : | : | | | : | : | | | : | : | | | : | : | | | : | : | | | : | : |
Db     63  NGKIFSDALEPDQETTYGVQVVFECNSGYMLDGPKIHCASAGGWSAETPKCFEIFCKPP 122
       1 : | : | | | : | : | | | : | : | | | : | : | | | : | : | | | : | : |
Qy    196  DVINGSPISQK 206
       : | | : |
Db    123  VILNGQAVLPK 133

```

```

RESULT      4
Q91275      PRELIMINARY;      PRT;      1053 AA.
AC      Q91275;
DT      01-NOV-1996 (TrEMBLrel. 01, Created)
DT      01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT      01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE      COMPLEMENT REGULATORY PLASMA PROTEIN.
OS      Parabrax nebulifer (barred sand bass).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC      Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC      Serranidae; Parabrax.
OX      NCBI_TaxID=30873;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=LIVER;
RX      MEDLINE=94318039; PubMed=8042982;
RA      Dahmen A., Kaidoh T., Ziefel P.F., Gigli I.;
RT      "Cloning and characterization of a cDNA representing a putative
RT      complement-regulatory plasma protein from barred sand bass (Parabrax
RT      nebulifer).";
RL      Biochem. J. 301:391-397(1994).
RL      EMBL; L21703; AAA92556.1; -
DR      HSP; P08603; LHFH
DR      InterPro; IPR000436; Sushi_SCR_CCP.
DR      Pfam; PF00084; sushi; 16.
DR      SMART; SM00032; CCP; 16.
SQ      SEQUENCE      1053 AA;      117597 MW;      F27E32C3AD76D5D3 CRC64;

```

Query Match	31.7%	Score 368.5;	DB 13;	Length 1053;
Best Local Similarity	39.6%;	Pred. No. 6.4e-29;		
Matches	72; Conservative	25; Mismatches	76; Indels	9; Gaps
QY	21	DQTPEGTQAIYKCRPGVRSIGNVMWCKRGEWALNPLRKCQRKPGHCGDTPFGFTL	80	
		: : : : :		
Db	47	EASIPGGRQVRVGCNVGS--GFPKLVCEGWETRQ--AKCQPSRCGHGDAQFADFHL	102	
		: : : : :		
QY	81	TGNVFVEYGKAVTYCNEGYQLLGEINRYECDDTDGWTNDIPCWBVKCLPVPAPENGKIV	140	
		: : : : : :		
Db	103	AEGNDFEFGSKVVYTCQKGQMVSRINYRCVAESGDGVVPVCESQCPLIHDNNVQVI	162	
		: : : : :		
QY	141	SSAMEPDREYHFQAQRFVNCNSGYKI--EGDEEMHCSDDGFWMSKEPKVCISCKSPDIV	199	
		: : : : :		
Db	163	GG-----PEATFNGVNFVRSCKSRSEILDGSPLEYCDRGDSWGSGVPVKCKAITCAIPPIEN	218	
		: : : : :		

QY 200 GS 201
 Db 219 GN 220

RESULT 5

Q9WRU2 PRELIMINARY; PRT; 645 AA.
 AC Q9WRU2;
 DT 01-NOV-1999 (TRENBLrel. 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE COMPLEMENT BINDING PROTEIN.
 OS Macaca mulatta rhadinovirus 17577.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=83534;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99174001; PubMed=100741154;
 RA Searles R.P., Bergquam E.P., Axthelm M.K., Wong S.W.;
 RT "Sequence and genomic analysis of a rhesus macaque rhadinovirus with
 RT similarity to Kaposi's sarcoma-associated Herpesvirus/Human
 RT herpesvirus 8.";
 RL J. Virol. 73:3040-3053(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Searles R.P., Bergquam E.P., Axthelm M.K., Wong S.W.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF083501; AAD21332.1;
 DR HSP: P10998; 1VVD.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; sushi; 8.
 DR SMART; SM00032; CCP; 8.
 SQ SEQUENCE 645 AA; 71526 MW; 93D8DE35ABF61EB2 CRC64;

Query Match 23.8%; Score 276.5; DB 12; Length 645;
 Best Local Similarity 31.1%; Pred. No. 9.1e-20;
 Matches 65; Conservative 35; Mismatches 90; Indels 19; Gaps 9;
 QY 1 EDCNELPPRRNTEITLGSMSDOTYPEGTQAIYKCRPGYRSLGNVIMV--CRKGEWVALNP 58
 Db 21 ENCK--PPHFEYRVKSNTEKOLYSVGETAELICRPGYVNTKIITTECLQGTW--STP 76
 QY 59 LRKCKRCPGCHGDPFGFTLTGG--NVFEYGVKAVYTCNEGQYLLGEINREC---DTD 114
 Db 77 NFPCKRCKPTPADLLNGAVHGGDNALKFGSNISYECNEGVDLIGS-NVRFCILQDTE 135
 QY 115 --GWTNDIPICEVVKCLPVTAPENGKIVSSAMEPDRE-YHFGQAVRFVCSNGYKIEGDEE 171
 Db 136 NVNWSNEPVCBIQCKP2PAVEHGDYL-----PNQDVYNGDAITFKCSLSYTLVGSTT 190
 QY 172 MHCSDGDFW-SKEKPKCIVEISCKSPDVING 200
 Db 191 LVCTSNKKNWSNFPCTCLMVCPSQIDNG 219

RESULT 6

Q9YTQ8 PRELIMINARY; PRT; 360 AA.
 AC Q9YTQ8;
 DT 01-MAY-1999 (TRENBLrel. 10, Created)
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE COMPLEMENT CONTROL PROTEIN HOMOLOG CCPH.
 OS Ateline herpesvirus 3.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=85618;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=73;
 RX MEDLINE=20091363; PubMed=10623770;
 RA Albrecht J.C.;
 RT "Primary structure of the Herpesvirus Ateles genome.";
 RL J. Virol. 74:1033-1037(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=73;
 RA Albrecht J.-C., Fleckenstein B.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF083424; AAC95530.1;
 DR HSP: P10998; 1VVD.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; sushi; 4.
 DR SMART; SM00032; CCP; 4.
 SQ SEQUENCE 360 AA; 40208 MW; 118CF83C034352A0 CRC64;

Query Match 23.6%; Score 275; DB 12; Length 360;
 Best Local Similarity 37.1%; Pred. No. 6.3e-20;
 Matches 78; Conservative 20; Mismatches 80; Indels 32; Gaps 14;
 QY 8 PRENTEI-----LTGSWSDOTYPEGTQAIYKCRPGYRSLGNVI--MVCRKGEWVALNPL 59
 Db 24 PKHRYVSLRYVNITN--SSGSYPNGTTLQVTCRKG--IGROIQVTCVNGNMTVPN-- 77
 QY 60 RKQKRCPCGHPGDTPTGFTLTGTVNFYGVKAVYTCNEGQYLLGEINREC--DTDG-- 115
 Db 78 -EQKRCRCSTPADLLNGWTYVT--GNLY-YGSVITYTCNTGYLLGSPT--SSCLLGPDCRV 133
 QY 116 -WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDREYH-FQGAQVRFVCSNGYK--IEGDEE 171
 Db 134 NMTTPRPPICEITKCKPPPTIANGHTNI-----KEYVTYLDVAVTSCNDETCLTLTGPS 188
 QY 172 MHCSDGDFW-SKEKPKCIVEISCKSPDVING 200
 Db 189 KQCSGTGRWVDEETKCEPKVKIPQVANG 218

RESULT 7

P88903 PRELIMINARY; PRT; 550 AA.
 AC P88903;
 DT 01-MAY-1997 (TRENBLrel. 03, Created)
 DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE ORF 4.
 OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=37296;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97094384; PubMed=8939871;
 RA Moore P.S., Boshoff C., Weiss R.A., Chang Y.;
 RT "Molecular mimicry of human cytokine and cytokine response pathway
 RT genes by KSHV.";
 RL Science 274:1739-1744(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97121480; PubMed=8962146;
 RA Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,
 RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
 RT "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus
 RT (HHV8).";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,
 RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U75698; AAC57082.1;
 DR HSP: P10998; 1VVD.

```
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 4.
DR SMART: SM00032; CCP; 4.
SQ SEQUENCE 550 AA; 60688 MW; D4B8B2B4BACD1CB5 CRC64;

Query Match 23.3%; Score 270.5; DB 12; Length 550;
Best Local Similarity 32.8%; Pred. No. 3.1e-19;
Matches 59; Conservative 24; Mismatches 80; Indels 17; Gaps 7;

QY 33 KCRPGYRSLG-NVIMVC-RKGEWALNPLRKCQKRCGHPGDPFGFTLTGG-NVFEYG 89
Db 52 RCRSGYTYARNITATCLQGWT--SEPTATCNKSKCPNPFGEIQNGKVIHFHGGQDALKYG 109

QY 90 VKAVYTCNEGYQLLGEINREC-----DTDGWTNDIPICEVVKCLPVTAPENCKIVSSAM 144
Db 110 ANISYVCNEGYFLVGRYVRYCMIGASGQMAWSSPPFCEKEC-----HRPKIENGDF 163

QY 145 EPDREYH-FQGAQVRFVNCYSYKIEGDEMHCHSDGFWKSEKPKCVCISCKSPDVINGSPI 203
Db 164 FKPKDYEYNDVAFHFCNEGYTLVGHPSIACAVNNTWTSNMPTCELACGCKFPSTVTHGYPI 223

RESULT 8
ID O40912 PRELIMINARY; PRT; 550 AA.
AC O40912;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ORF 04.
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=37296;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97296220; PubMed=9151804;
RA Neipel F., Albrecht J.C., Fleckenstein B.;
RT "Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus
human herpesvirus 8: determinants of its pathogenicity?";
RL J. Virol. 71:4187-4192(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Neipel F., Albrecht J.-C., Ensser A., Huang Y.-Q., Li J.J.,
RA Friedman-Kien A.E., Fleckenstein B.;
RT "The genome of human herpesvirus 8 cloned from Kaposi's sarcoma.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: U93872; AAB82602.1; -.
DR HSSP: P10998; IVD.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 4.
DR SMART: SM00032; CCP; 4.
SQ SEQUENCE 550 AA; 60648 MW; 3A7FC1C30F79C6A6 CRC64;

Query Match 23.3%; Score 270.5; DB 12; Length 550;
Best Local Similarity 32.8%; Pred. No. 3.1e-19;
Matches 59; Conservative 24; Mismatches 80; Indels 17; Gaps 7;

QY 33 KCRPGYRSLG-NVIMVC-RKGEWALNPLRKCQKRCGHPGDPFGFTLTGG-NVFEYG 89
Db 52 RCRSGYTYARNITATCLQGWT--SEPTATCNKSKCPNPFGEIQNGKVIHFHGGQDALKYG 109

QY 90 VKAVYTCNEGYQLLGEINREC-----DTDGWTNDIPICEVVKCLPVTAPENCKIVSSAM 144
Db 110 ANISYVCNEGYFLVGRYVRYCMIGASGQMAWSSPPFCEKEC-----HRPKIENGDF 163

QY 145 EPDREYH-FQGAQVRFVNCYSYKIEGDEMHCHSDGFWKSEKPKCVCISCKSPDVINGSPI 203
Db 164 FKPKDYEYNDVAFHFCNEGYTLVGHPSIACAVNNTWTSNMPTCELACGCKFPSTVTHGYPI 223

RESULT 9
ID Q9J2M6 PRELIMINARY; PRT; 395 AA.
AC Q9J2M6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE COMPLEMENT BINDING PROTEIN.
OS Macaca mulatta rhadinovirus 26-95.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=119193;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MACACA MULATTA RHADINOVIRUS ISOLATE 26-95;
RX MEDLINE=20173730; PubMed=10708456;
RA Alexander L., Denekamp L., Knapp A., Auerbach M.R., Damania B.,
RA Desrosiers R.C.;
RT "The primary sequence of rhesus monkey rhadinovirus isolate 26-95:
sequence similarities to Kaposi's sarcoma-associated herpesvirus and
rhesus monkey rhadinovirus isolate 17577.";
RL J. Virol. 74:3388-3398(2000).
DR EMBL: AF210726; AAF59982.1; -.
DR HSSP: P10998; IVD.
DR InterPro: IPR001230; Prenyltn.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 4.
DR SMART: SM00032; CCP; 4.
DR PROSITE: PS00294; PRENYLATION; UNKNOWN_1.
SQ SEQUENCE 395 AA; 43922 MW; B4C9C6F2E226AE06 CRC64;

Query Match 23.2%; Score 269.5; DB 12; Length 395;
Best Local Similarity 31.7%; Pred. No. 2.6e-19;
Matches 65; Conservative 34; Mismatches 87; Indels 19; Gaps 10;

QY 7 PPRNTEILTGSMSDQTYPEGTQAIYKCRPG-YRSLGNVIMVC-RKGEWALNPLRKCQK 64
Db 30 PPEDRFVWKAN-QENYAVGTRVELICRPGYKQANVYVECLSNGTWT--TPNAECRR 86

QY 65 RCGHPCDTPFGFTLT-GGVNFEYGVKAVTTCNEGYQLLGEINREC-----DTD--GWT 117
Db 87 KRCSNPEDILNGEVIITDSNAPKFGSNITYKNTGTVLLLG-ATVTRTCLLYKYSNLVDWQ 145

QY 118 NDIPICEVVKCLPVTAPENCKIVSSAMEPDRE-YHFGQAVRFVNCYSYKIEGDEMHCHSD 176
Db 146 PAAPTCEIECKKQPDIECKYY-----PVQEFYNYLETITFTCNKDFSLIGNTTTTCMT 200

QY 177 DGFWSKEKPKCVCISCKSPDVINGS 201
Db 201 NGTWSSPVKPCQITCSAPNIDHGT 225

RESULT 10
ID Q96QU9 PRELIMINARY; PRT; 3389 AA.
AC Q96QU9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CUB AND SUSHI MULTIPLE DOMAINS PROTEIN 1 SHORT FORM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21365705; PubMed=11472063;
RA Sun P.C., Uppaluri R., Schmidt A.P., Pashla M.E., Quant E.C.,
RA Sunwoo J.B., Gollin S.M., Scholnick S.B.;
RT "Transcript map of the 8p23 putative tumor suppressor region.";
RL Genomics 75:17-25(2001).
DR EMBL: AY017307; AAG52948.1; -.

```


DR	InterPro: IPR000436; Sushi_SCR_CCP.
DR	InterPro: IPR002035; WMFA.
DR	Pfam: PF00008; EGF; 10.
DR	Pfam: PF02494; HYR; 2.
DR	Pfam: PF00084; sushi; 33.
DR	Pfam: PF00092; vwa; 1.
DR	PRINTS: PR00010; EGFBL00D.
DR	PRINTS: PR00895; PENTAXIN.
DR	PRINTS: PR00453; VWFADOMAIN.
DR	ProDom: PD002153; Pentaxin; 1.
DR	SMART: SM00032; CCP; 34.
DR	SMART: SM00181; EGF; 15.
DR	SMART: SM00179; EGF_CA; 9.
DR	SMART: SM00001; EGF_like; 3.
DR	SMART: SM00159; PTX; 1.
DR	SMART: SM00327; VWA; 1.
DR	PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR	PROSITE: PS00234; WMFA; 1.
KW	Signal.
FT	SIGNAL
ST	1
ST	17
ST	POTENTIAL.
ST	SEQUENCE 3567 AA; 387391 MW; 8FBA8276E12293E5 CRC64;
ST	Query Match 22.4%; Score 260; DB 11; Length 3567;
ST	Best Local Similarity 29.1%; Pred. No. 3.6e-17;
ST	Matches 55; Conservative 30; Mismatches 92; Indels 12; Gaps 4;
QY	17 GSWSDQTYPEQTQAIYKCRPGYRSLGNVIVC-RKGEWALNPLRKCKQKPCGHPGDTPF 75
DB	1797 GHSSGEIYVGTAVTFSCDEGHELVGYSTICTLETGWDLRP--SCEAISCGVP---PV 1851
QY	76 GTFTLTGCVNFYGVKAVYTCNBYQYLLGEINRECDTGTNDIPICEVVKCLPVTAPE 135
DB	1852 PENGGVGDGSAFTYGSKYVYRCDKGYTSLGDSGDESACLASGWSHSSPYCELVKCSQPEDIN 1911
QY	136 NGKTVSSAMEPDREYHFQAVRFVCSNGYSKIEGDEEMHCSDDGWPSKEKPKCVIEICKSP 195
DB	1912 NGKYLISGLT-----YLSIASYSCENGYSLQGPSLLECTAGSGWDRAPPSQQLVSCGEP 1965
QY	196 DVGINSPIIS 204
DB	1966 PIVKDAVIT 1974
RESULT	15
Q923L3	PRELIMINARY; PRT; 3564 AA.
ID	Q923L3
AC	Q923L3;
DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE	CSMD1.
GN	CSMD1.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6;
RA	MEDLINE=21365705; PubMed=11472063;
RA	Sun P.C., Uppaluri R., Schmidt A.P., Pashia M.E., Quant E.C.,
RA	Sunwoo J.B., Gollin S.M., Scholnick S.B.;
RT	"Transcript map of the 8p23 putative tumor suppressor region.";
RL	Genomics 75;17-25(2001).
RL	EMBL: AK01475; AAG54083.1; ..
ST	SEQUENCE 3564 AA; 387865 MW; 70824C55B0674609 CRC64;
ST	Query Match 22.2%; Score 258.5; DB 11; Length 3564;
ST	Best Local Similarity 30.0%; Pred. No. 5.1e-17;
ST	Matches 57; Conservative 32; Mismatches 84; Indels 17; Gaps 6;

QY 14 ILTGSWSDQTYPEGTOAIYKCRPGYRSLGNVIMVC-----RKGEWVALNPLRKCQKRPCG 68
Db 2686 IVNGHISGDGFSYRDTWVQC�PFRVLGTSVRICCTTSGRGR-----LTVCVPITCG 2739
QY 69 HPGDTPFGFTLTGGNVFEYGVKAVYTCNEGYOLLGEINRYRECDTDG-WTNDIPICEVVK 127
Db 2740 HPGNPAHG---LTNGTEFNLDLVNFTCHTGYRLQG-ASRAQCRSNGQWSSPLPICRVVN 2795
QY 128 CLPVTAPENCKIVSSAMEPDREYHFGQAVFVCNCGYKIEGDEEMHCSDDGFWSEKPKC 187
Db 2796 CSDPGSVENAVRHGQONFPE-SFEYGTŠVNYHCKTGfYLLIGSSALTCMASGLWDRSLPKC 2854
QY 188 VEISCKSPDV 197
Db 2855 LAISCGHPGV 2864

Search completed: August 29, 2002, 15:11:38
Job time: 599 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:00:49 ; Search time 75.55 Seconds
(without alignments)
304.332 Million cell updates/sec

Title: US-09-316-163-9

Perfect score: 1163

Sequence: 1 EDCELPPRNTTEILGWS.....VEISCKSPDVINGSPISOKI 207

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	721	62.0	240	18 AAW39154	Human partial Comp
2	627	53.9	216	18 AAW39155	Clone PRB9PH410 C
3	276.5	23.8	645	21 AAB53125	Macaca mulatta rha
4	261.5	22.5	613	22 AAU00816	Human Immunoglobul
5	261.5	22.5	882	22 AAB83372	NOV16 protein sequ
6	261	22.4	299	17 AAW06881	Decay accelerating
7	261	22.4	376	20 AAY50035	Human complement r
8	261	22.4	381	8 AAP70048	Human decay accele
9	261	22.4	381	10 AAP94773	Decay accelerating
10	261	22.4	381	16 AAR66683	Decay accelerating
11	261	22.4	381	18 AAW26317	Human decay accele

12	261	22.4	381	20 AAY31740	Human CD55 and 791
13	261	22.4	381	20 AAW73505	Decay accelerating
14	261	22.4	440	8 AAP70049	Human decay accele
15	261	22.4	440	10 AAP94774	Membrane bound dec
16	261	22.4	440	16 AAR66684	Decay accelerating
17	261	22.4	440	18 AAW27483	Human glycophospha
18	261	22.4	543	13 AAR28557	CRI-4 (99H, 103E)
19	261	22.4	577	17 AAE06882	Membrane co-factor
20	261	22.4	611	22 AAE12569	CAB2 protein. Uni
21	261	22.4	611	22 AAE03762	CAB-2 chimeric pro
22	259	22.3	497	22 ABB10508	Human CDNA SEQ ID
23	259	22.3	497	22 AAU18149	Novel human uterin
24	259	22.3	497	22 AAU17035	Human novel secret
25	259	22.3	497	22 AAU19960	Novel human calciu
26	259	22.3	515	22 AAG93953	Human polypeptide,
27	259	22.3	543	13 AAR28547	CRI-4 (52S, 53S, 5
28	259	22.3	1139	22 ABB10326	Human CDNA SEQ ID
29	259	22.3	1139	22 AAU18126	Novel human uterin
30	259	22.3	1139	22 AAU16963	Human novel secret
31	259	22.3	1139	22 AAU19902	Novel human calciu
32	257.5	22.1	496	20 AAY5752	Human CRI protein
33	257.5	22.1	543	13 AAR28560	CRI-4 (114S) analo
34	256.5	22.1	263	22 AAB48846	Cowpox virus infla
35	255.5	22.0	320	22 AAG68150	Codon modified hum
36	255	21.9	254	15 AAR47154	Sequence of solubi
37	255	21.9	254	15 AAR47155	CRI-4 (64K) analog
38	255	21.9	543	13 AAR28550	CRI-4 (85R, 87N) a
39	255	21.9	543	13 AAR28553	CRI-4 (85R, 87N) a
40	255	21.9	543	13 AAR28565	CRI-4 (121Q) analo
41	255	21.9	543	13 AAR28566	CRI-4 (318R, 319N)
42	255	21.9	543	13 AAR28567	CRI-4 (318-321 RNP
43	255	21.9	543	13 AAR28568	CRI-4 (347T, 349Y)
44	255	21.9	543	13 AAR28569	CRI-4 (369-376 STK
45	255	21.9	543	13 AAR28570	CRI-4 (266-274 KLK

ALIGNMENTS

RESULT 1

AAW39154

ID AAW39154 standard; Protein; 240 AA.

XX AAW39154;

XX

XX

XX 27-APR-1998 (first entry)

DT Human partial Complement factor H protein fragment 1.

DE

DE

XX Complement factor H; tumour associated antigen; renal cancer;

KW urogenital cancer; medicament; modulator.

XX

OS Homo sapiens.

XX

XX WO9738136-A1.

XX

PD 16-OCT-1997.

XX

XX

PF 09-APR-1997; 97WO-US05710.

XX

PR 06-MAR-1997; 97US-0812481.

PR 09-APR-1996; 96US-0015083.

PR 09-APR-1996; 96US-0630048.

PR 06-MAR-1997; 97US-0038614.

(BARD-) BARD DIAGNOSTIC SCI INC.

PI Enfield DL, Hass GM, Kinders RJ;

DR WPI; 1997-512742/47.

DR N-PSDB; AAV02790.

XX

PT Treating or screening for cancer, e.g. renal or urogenital cancer -

PT by modulating or detecting tumour associated human complement Factor
PT H related antigen, or nucleic acid encoding it
XX Example 6B; Fig 6B; 104pp; English.
XX
CC This partial protein sequence represents a region of the human
CC tumour-associated complement factor H (CFH). This sequence is used
CC in the identification of complement factor H related proteins and
CC antigens isolated from clone pRRB9FH410 (see AAW39155). The detection of
CC such proteins and a CFH antigens can be used in screening or for the
CC treatment of renal or urogenital cancer, e.g. bladder, cervical or
CC prostate cancer. Agents that may modulate this antigen could be used in
CC the manufacture of a medicament for the treatment of a tumour cell.
XX
SQ Sequence 240 AA;

Query Match 62.0%; Score 721; DB 18; Length 240;
Best Local Similarity 100.0%; Pred. No. 3.5e-59;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 FTLTGNNVFEGYGVKAVTCTNEGQYLLGEINRYRECDDTGDWTDIPICEVVKCLPVTAPENG 137
|||||
Db 1 flltgnvfeygvkavtctnegyqllgeinryrecddtgdwtndipicevkvclpvtapeng 60

QY 138 KIVSSAMEPDREYHFGQAVRVCNSGYKIEGDEEMHCDDGFWSKPKCIVEISCKSPDV 197
|||||
Db 61 kivssamepdrehyfkgavrfvcnsgykiedgemhcsddgfwskpkcveisckspdv 120

QY 198 INGSPTISQKI 207
|||||
Db 121 ingspisqki 130

RESULT 2
AAK39155
ID AAW39155 standard; Protein; 216 AA.
XX
AC AAW39155;
XX
DT 27-APR-1998 (first entry)
XX
DE Clone pRRB9FH410 CFH related protein fragment.
XX
KW Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medicament; modulator.
XX
OS Synthetic.
XX
PN WO9738136-A1.
XX
PD 16-OCT-1997.
XX
PF 09-APR-1997; 97WO-US05710.
XX
PR 06-MAR-1997; 97US-0812481.
PR 09-APR-1996; 96US-0015083.
PR 09-APR-1996; 96US-0630048.
PR 06-MAR-1997; 97US-0038614.
XX
PA (BARD-) BARD DIAGNOSTIC SCI INC.
XX
PI Enfield DL, Hass GM, Kinders RJ;
XX
DR WPI; 1997-512742/47.
DR N-PSDB; AAV02791.
XX
XX Treating or screening for cancer, e.g. renal or urogenital cancer -
PT by modulating or detecting tumour associated human complement factor -
PT H related antigen, or nucleic acid encoding it
XX
PS Example 6B; Fig 6B; 104pp; English.
XX

CC This partial protein is found in clone pRRB9FH410 and represents a
CC complement factor H related protein with homology to a region of the
CC human tumour-associated complement factor H (CFH). The detection of this
CC protein and a CFH antigen can be used in screening or for the treatment
CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.
CC Agents that may modulate this antigen could be used in the manufacture of
CC a medicament for the treatment of a tumour cell.
XX
SQ Sequence 216 AA;

Query Match 53.9%; Score 627; DB 18; Length 216;
Best Local Similarity 99.1%; Pred. No. 1.7e-50;
Matches 112; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 95 TCNEGYQLLGEINRYRECDDTGDWTDIPICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQ 154
|||||
Db 1 tcnegyqllgeinryrecddtgdwtndipicevkvclpvtapengkivssamepdrehyfqq 60

QY 155 AVRFVCSNGYKIEGDEEMHCDDGFWSKPKCIVEISCKSPDVINGSPISQKI 207
|||||
Db 61 avrfvcnsgykiedgemhcsddgfwgkpkcveisckspdvingspisqki 113

RESULT 3
AAB53125
ID AAB53125 standard; Protein; 645 AA.
XX
AC AAB53125;
XX
DT 28-FEB-2001 (first entry)
XX
DE Macaca mulatta rhadinovirus 17577 RRV ORF4 protein SEQ ID NO:7.
XX
KW Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;
KW genome; Kaposi's sarcoma-associated herpesvirus; KSHV; Interleukin 6;
KW IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;
KW cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;
KW lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
KW splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia.
XX
OS Macaca mulatta rhadinovirus 17577.
XX
PN WO200028040-A2.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26260.
XX
PR 06-NOV-1998; 98US-0107507.
PR 20-NOV-1998; 98US-0109409.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Wong SW, Axthelm MK, Searles RP;
XX
DR WPI; 2000-376552/32.
XX
PT New rhesus rhadino virus for producing non-human primate model useful
PT for testing potential treatments and efficacy of the candidate vaccine
PT for conditions associated with RRV infection -
XX
PS Claim 5; Page 122-123; 141pp; English.
XX
CC The present invention describes a novel rhesus macaque rhadinovirus
CC called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the
CC RRV genome sequence, and AAB53123 to AAB53204 represent the proteins
CC encoded by the genome sequence. The present invention also specifically
CC claims the individual open reading frame (ORF) nucleotide sequences from
CC the genome which encode the individual proteins, but these sequences are
CC not given. A non-human animal infected with RRV can be used for testing
CC the efficacy of drug in the treatment of condition associated with
CC infection with RRV such as Kaposi's sarcoma, lymphoproliferative

PH	Key	Location/Qualifiers
FT	Domain	4..52
FT		/label= Immunoglobulin_like_domain
FT	Region	6..11
FT		/label= Immunogenic_epitope
FT	Region	73..79
FT		/label= Immunogenic_epitope
FT	Region	121..137
FT		/label= Immunogenic_epitope
FT	Region	141..146
FT		/label= Immunogenic_epitope

Qy	14	ILTGWSDDQYPEGTOAIYKCRPGYRSLGNIMVCRKG - EWVALNPLRKCKQRPCGHPGD	72
Db	217	lvnghtngenysrgsvvqcnagflqlmsrviccadhhwsktbf--cvaitcqhpan	274

PT comprise membrane co-factor protein and decay accelerating factor
PT peptide sequences
XX
PS Disclosure; Page 25; 33pp; English.
XX
CC A portion (AAW06881) of the complement-inhibitor, decay accelerating
CC factor (DAF), is used in novel chimeric proteins of formula
CC A-R1-B-R2-C, where A and C are peptides (AAW06875-79, AAW06883-90) able
CC to bind glycosaminoglycans (esp. heparin) present on cell surfaces,
CC R1 is a portion of DAF or membrane co-factor protein (MCP, see also
CC AAW06880), R2 is DAF when R1 is MCP or MCP when R1 is DAF, and B is a
CC peptide that may have complement inhibitor activity. The chimeric
CC proteins (see also AAW06882) are directed to cell surfaces where they
CC inhibit complement-mediated cell lysis. They are used to treat and
CC prevent disease states in which complement plays a role, e.g.
CC sepsis, adult respiratory distress syndrome, reperfusion injury and
CC tissue damage.
XX
SQ Sequence 299 AA;

Query Match 22.4%; Score 261; DB 17; Length 299;
Best Local Similarity 29.9%; Pred. No. 3.1e-16;
Matches 70; Conservative 42; Mismatches 84; Indels 38; Gaps 12;

QY 1 EDCN---ELPPRRNTEILTGSMSDQTY-PEGTQAIYKCRPGYR---SLGNVIMVCRKGEW 53
Db 34 efncrscevprrlnsaslkqpyitqnyfpvgtvveyecrpgyrrrepslspkltclqlnkw 93
QY 54 VALNPLRCKRCPGHPGDTPTFTLTGNVFEYGVKAVYTCNEGYSQLLGEINYRECDT 113
Db 94 --stavefckkscnpgeirngqidvpqglf--gatisfscntgylfgsts-sfc11 148
QY 114 DG-----WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDRE-YHFGQAVRFVCSGKIEG 168
Db 149 sgssvqwsdplpecrelyc-----pappqidngliqgerdhgygsvtyacnkgftmlg 203
QY 169 DEEMHC---SDDGFSWSEKPKC-----VEISCKSPDVIN-----GSPISQK 206
Db 204 ehsiyctvnndegewsgpppecrgksltskvpptvgkpttnvpttsvtsqk 257

RESULT 7
ID AAY50035
XX AAY50035 standard; Protein; 376 AA.
AC AAY50035;
XX
DT 19-JAN-2000 (first entry)
XX
DE Human complement regulatory protein DAF.
XX
KW Complement regulatory protein; decay accelerating factor; DAF;
KW immune reaction; xenotransplantation; xenograft; transplant; organ;
KW rejection; hyperacute; inhibition; protection; heart; lung; liver;
KW kidney; pancreas; thyroid; islet cell; neurons; stem cell; tissue;
KW skin.
XX
OS Homo sapiens.
XX
XX W09953042-A2.
XX
XX 21-OCT-1999.
XX
XX 08-APR-1999; 99WO-GB01085.
XX
XX 09-APR-1998; 98GB-0007520.
XX
XX (UYWA-) UNIV WALES COLLEGE OF MEDICINE.
XX
XX Morgan BP, Rushmere NK, Hinchliffe SJ, Van Den Berg CW;
XX WPI; 1999-620420/53.
XX

XX Use of cells or tissues expressing complement regulatory molecules for,
PT e.g. preventing xenotransplant rejection in humans -
XX
XX Claim 18; Fig 15; 88pp; English.
XX
CC This sequence represents human complement regulatory protein DAF
CC (decay accelerating factor). Complement regulatory proteins (CRPs)
CC such as DAF and CD59 are membrane bound proteins that protect an
CC organism's cells from attack by its own complement. Hyperacute
CC rejection of a xenotransplant occurs because natural antibodies in
CC the human recipient bind to the endothelium of the donor organ and
CC activate complement, thereby initiating rapid rejection.
CC Hyperexpression of this protein in pig cells may be useful for the
CC protection of xenotransplanted organs, as greatly increased amounts
CC of functional CRP molecules will be expressed. The organ should be
CC resistant to attack by human complement, thus preventing organ
CC rejection. Methods of causing donor animal cells to hyperexpress CRPs
CC can be used for protection of xenotransplants. For example, organs,
CC tissue and cells can be generated which are resistant to complement
CC attack and hence to hyperacute rejection when transplanted into humans.
CC The methods can be used for organs, such as heart, lung, liver, kidney,
CC pancreas and thyroid; cells, such as islet cells, neurons, and stem
CC cells; or tissues, such as skin.
XX
SQ Sequence 376 AA;

Query Match 22.4%; Score 261; DB 20; Length 376;
Best Local Similarity 29.9%; Pred. No. 4e-16;
Matches 70; Conservative 42; Mismatches 84; Indels 38; Gaps 12;

QY 1 EDCN---ELPPRRNTEILTGSMSDQTY-PEGTQAIYKCRPGYR---SLGNVIMVCRKGEW 53
Db 87 efncrscevprrlnsaslkqpyitqnyfpvgtvveyecrpgyrrrepslspkltclqlnkw 146
QY 54 VALNPLRCKRCPGHPGDTPTFTLTGNVFEYGVKAVYTCNEGYSQLLGEINYRECDT 113
Db 147 --stavefckkscnpgeirngqidvpqglf--gatisfscntgylfgsts-sfc11 201
QY 114 DG-----WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDRE-YHFGQAVRFVCSGKIEG 168
Db 202 sgssvqwsdplpecrelyc-----pappqidngliqgerdhgygsvtyacnkgftmlg 256
QY 169 DEEMHC---SDDGFSWSEKPKC-----VEISCKSPDVIN-----GSPISQK 206
Db 257 ehsiyctvnndegewsgpppecrgksltskvpptvgkpttnvpttsvtsqk 310

RESULT 8
ID AAP70048
XX AAP70048 standard; protein; 381 AA.
AC AAP70048;
XX
DT 03-FEB-1991 (first entry)
XX
DE Human decay acceleration factor variant #1.
XX
XX Decay acceleration factor.
XX
XX Key Location/Qualifiers
XX 331..347
XX /label=putative transmembrane region
XX
XX EP244267-A.
XX
XX 04-NOV-1987.
XX
XX 01-MAY-1987; 87EP-0303944.
XX
XX 02-MAY-1986; 86US-0859107.
XX

PA (GETH) GENENTECH INC.
 XX Caras IW;
 XX WPI; 1987-308481/44.
 DR N-PSDB; AAN70047.
 XX New decay accelerating factor variants - obtained with the factor by
 PT using recombinant DNA procedures.
 XX Disclosure; Page 15-17; 20pp; English.
 XX The probable phosphatidylinositol derivatization site is Cys(330).
 CC The DAF variant is useful for treating paroxysmal nocturnal
 CC haemoglobinuria, or inflammatory or cell lytic autoimmune
 CC diseases. It may be used to ameliorate allograft rejection
 CC or autoimmune diseases. See also AAN70046, AAN70048.
 XX Sequence 381 AA;
 SQ

Query Match 22.4%; Score 261; DB 8; Length 381;
 Best Local Similarity 29.9%; Pred. No. 4.1e-16;
 Matches 70; Conservative 42; Mismatches 84; Indels 38; Gaps 12;

QY 1 EDCN---ELPPRRNTEILTGSWSDQTY-PEGTQAIYKCRPGYR---SLGNVIMVCRKGEW 53
 Db | | | | | : | : | | | | | | | | | | : | : |
 92 efncrscevptrlnasakqpyitqnyfpvgtvveecrpgyrrepslskltclqnlkw 151
 QY 54 VALNPLRKCKRCPGHPGDTPEFTLTGNNVFEGVKAAYTCNEGYQLLGEINRECDT 113
 Db : | | : | | | : | : | | | : | : | | | | : | : |
 152 --stavefckkscnpngeirngqidvpqglif--gatisfscntgyklfgsts-sfc11 206

QY 114 DG-----WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDRE-YHFGQAVRFVCSGYKIEG 168
 Db | | | : | | : | : | : | : | : | : | : | : | : | : |
 207 sgssvqwsdp1pecreiyc-----pappqidingliqgerdhgyrqsvtacnkgftmig 261

QY 169 DEEMHC---SDGDFWSKEKPKC-----VEISCKSPDVIN-----GSPISQK 206
 Db : | : | : | | | | | | : | : | : | : | : | : | | | | |

RESULT 9
 AAP94773
 ID AAP94773 standard; protein; 381 AA.
 AC AAP94773;
 XX 04-JUL-1990 (first entry)
 XX Decay accelerating factor (DAF) of clones lambda 33 and lambda 47.
 DE DAF; allograft rejection; affinity purification;
 XX autoimmune disease; ds.
 OS Synthetic.
 XX W08901041-A.
 PN 09-FEB-1989.
 XX 03-AUG-1988; 88WO-US02648.
 PR 06-AUG-1987; 87US-0083757.
 XX (GETH) GENETECH INC.
 PA Caras I;
 PI WPI; 1989-061177/08.
 DR N-PSDB; AAN91043.
 XX Fusion polypeptide for targeting protein to cell membrane -
 PT

PT comprisesphospholipid anchor domain with heterologous
 PT polypeptide.
 XX Disclosure; ; 6lpp; English.
 PS Recombinant DAF's are useful in treatment of inflammatory or cell lytic
 XX autoimmune diseases and allograft rejection. Useful in diagnostic
 CC compositions or in affinity purification.
 XX Sequence 381 AA;
 SQ

Query Match 22.4%; Score 261; DB 10; Length 381;
 Best Local Similarity 29.9%; Pred. No. 4.1e-16;
 Matches 70; Conservative 42; Mismatches 84; Indels 38; Gaps 12;

QY 1 EDCN---ELPPRRNTEILTGSWSDQTY-PEGTQAIYKCRPGYR---SLGNVIMVCRKGEW 53
 Db | | | | | : | : | | | | | | | | | | : | : |
 92 efncrscevptrlnasakqpyitqnyfpvgtvveecrpgyrrepslskltclqnlkw 151

QY 54 VALNPLRKCKRCPGHPGDTPEFTLTGNNVFEGVKAAYTCNEGYQLLGEINRECDT 113
 Db : | | : | | | : | : | | | : | : | | | | : | : |
 152 --stavefckkscnpngeirngqidvpqglif--gatisfscntgyklfgsts-sfc11 206

QY 114 DG-----WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDRE-YHFGQAVRFVCSGYKIEG 168
 Db | | | : | | : | : | : | : | : | : | : | : | : | : |
 207 sgssvqwsdp1pecreiyc-----pappqidingliqgerdhgyrqsvtacnkgftmig 261

QY 169 DEEMHC---SDGDFWSKEKPKC-----VEISCKSPDVIN-----GSPISQK 206
 Db : | : | : | | | | | | : | : | : | : | : | : | | | | |

RESULT 10
 AAR66683
 ID AAR66683 standard; Protein; 381 AA.
 AC AAR66683;
 XX 23-JUL-1995 (first entry)
 XX Decay accelerating factor.
 DE Decay accelerating factor; DAF; mDAF; fusion protein; liposome;
 KW cell targeting; glycoposphatidylinositol; GPI; drug delivery.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FT Peptide 1..34
 FT /label= sig_peptide
 FT Modified-site 364
 FT /note= "probable phosphatidylinositol
 FT derivatization site".
 XX US5374548-A.
 XX 20-DEC-1994.
 XX 02-MAY-1986; 86US-0859107.
 XX 02-MAY-1986; 86US-0859107.
 PR 06-AUG-1987; 87US-0083757.
 PR 19-DEC-1991; 91US-0811048.
 PR 12-FEB-1993; 93US-0017934.
 XX (GETH) GENENTECH INC.
 PA Caras IW;
 PI WPI; 1995-035649/05.
 DR N-PSDB; AAR66683.
 XX

(ALEX-) ALEXION PHARM INC.
Mason JM, Squinto SP;
WPI; 1997-350243/32.
N-PSDB; AAT84471.

XX	WO9943800-A1.
PN	
XX	
PD	02-SEP-1999.
PD	
XX	
PF	26-FEB-1999;
XX	99WO-GB00582.
PR	26-FEB-1998;
	98GB-0004065.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2002, 06:04:46 ; Search time 58.13 Seconds
(without alignments)
438.047 Million cell updates/sec

Title: US-09-316-163-10

Perfect score: 1497

Sequence: 1 EDCELPPRRNTEILTGWS.....EKSCDNPYPNGDYSPLRIK 265

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR71:**

1: pir1:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1497	100.0	449	1 NBHUHS	complement factor
2	1497	100.0	1231	1 NBHSH	complement factor
3	1024	68.4	1234	1 NBKSH	complement factor
4	699	46.7	669	2 S65551	factor H - bovine
5	440	29.4	1053	2 S46199	probable complement
6	332.5	22.2	360	2 T42921	complement control
7	303	20.2	597	1 NBHUC4	C4b-binding protein
8	302	20.2	597	1 S53711	C4BP alpha chain p
9	301.5	20.1	263	1 WMVZSP	apolipoprotein H h
10	301	20.1	676	2 A45900	complement C3b rec
11	298.5	19.9	2489	2 I73012	complement C3b/C4b
12	290.5	19.4	2014	2 I36336	complement recepto
13	289.5	19.3	497	2 JC2054	complement regulat
14	284.5	19.0	263	1 C36838	complement control
15	284.5	19.0	263	2 T28450	hypothetical prote
16	284.5	19.0	482	2 A34924	complement C3b/C4b
17	283.5	18.9	263	2 B72152	B18L protein - var
18	277.5	18.5	560	2 T16833	hypothetical prote
19	276.5	18.5	558	2 S57953	C4BP protein alpha
20	275	18.4	469	1 NBMSC4	C4b-binding protei
21	274.5	18.3	349	2 G02913	sperm CD46 - human
22	274.5	18.3	369	2 I57998	membrane cofactor
23	274	18.3	377	2 I54479	membrane cofactor
24	274	18.3	384	2 S01896	membrane cofactor
25	272	18.2	610	2 A35046	E-selectin precurs
26	270.5	18.1	369	2 JC5194	membrane cofactor
27	270.5	18.1	362	2 JC5138	membrane cofactor
28	270.5	18.1	440	2 A43519	complement recepto
29	266	17.8	482	2 JC5092	E-selectin - pig

30	265.5	17.7	610	1 I46001	C4b-binding protei
31	264.5	17.7	302	1 WMBE1E	secretory compleme
32	264.5	17.7	360	1 WMBE2E	membrane-bound com
33	262	17.5	381	1 B26359	decay-accelerating
34	262	17.5	440	2 A26359	decay-accelerating
35	260.5	17.4	768	2 A42755	P-selectin precurs
36	260	17.4	612	2 B42755	complement C3d/Eps
37	257.5	17.2	1025	1 A43526	complement C3d/Eps
38	257	17.2	579	2 A56740	sperm-egg recognit
39	255	17.0	452	2 A35068	complement factor
40	254	17.0	768	2 I53821	P-selectin - rat
41	253	16.9	345	1 NBMS	apolipoprotein H p
42	252.5	16.9	1091	1 PL0009	complement C3d/Eps
43	252	16.8	340	2 I56234	decay-accelerating
44	252	16.8	661	1 KFHU13	coagulation factor
45	250.5	16.7	830	2 A30359	P-selectin precurs

ALIGNMENTS

RESULT 1

NBHUHS

complement factor H precursor, short splice form [validated] - human
N;Alternate names: complement factor H-related protein; complement protein H
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence.revision 23-Feb-1996 #text.change 08-Dec-2000
C;Accession: S03013; B60238; A27877; A61103; A26505; S10479
R;Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.

Biochem. J. 249, 593-602, 1988

A;Title: The complete amino acid sequence of human complement factor H.

A;Reference number: S00254; MUID:88134059

A;Accession: S03013

A;Molecule type: mRNA

A;Residues: 1-449 <RIP>

A;Cross-references: EMBL:X07523; EMBL:Y00716; NID:g32492; PIDN:CAA30403.1; PID:g75807

A;Note: Part of this sequence, including the amino end of the mature protein was conf

A;Note: 402-Tyr was also found

R;Estaller, C.; Schwaebel, W.; Stanley, K.K.; Weiss, E.; Dierich, M.P.

Eur. J. Immunol. 21, 799-802, 1991

A;Title: Human complement factor H: two factor H proteins are derived from alternativ

A;Reference number: A60238; MUID:91184292

A;Accession: B60238

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-33;434-449 <EST>

A;Note: only portions of this 1.8 kilobase mRNA were sequenced

R;Schulz, T.F.; Schwaebel, W.; Stanley, K.K.; Weiss, E.; Dierich, M.P.

Eur. J. Immunol. 16, 1351-1355, 1986

A;Title: Human complement factor H: isolation of cDNA clones and partial cDNA sequenc

A;Reference number: A27877; MUID:87054207

A;Accession: A27877

A;Molecule type: mRNA

A;Residues: 1-IL, 55-401, 'V', 403-449 <SCH>

A;Cross-references: GB:X04697; NID:g31991; PIDN:CAB41739.1; PID:g4725976

A;Note: an additional nucleotide present within the codon for Glu-310 was thought to

R;Schwaebel, W.; Zwirner, J.; Schulz, T.F.; Linke, R.P.; Dierich, M.P.; Weiss, E.H.

Eur. J. Immunol. 17, 1485-1489, 1987

A;Title: Human complement factor H: expression of an additional truncated gene produc

A;Reference number: A61103; MUID:88055295

A;Accession: A61103

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 27-76 <SC2>

A;Note: this is a partial sequence of an alternatively spliced 1.8 kilobase mRNA that

R;Sim, R.B.; Discipio, R.G.

Biochem. J. 205, 285-293, 1982

A;Title: Purification and structural studies on the complement-system control protein

A;Reference number: A26505; MUID:83048213

A;Accession: A26505

A;Molecule type: protein

A;Residues: 19-20, 'Q', 22-29, 'V', 31-33, 'Q', 35 <SIM>

R;Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P

Biochemistry 31, 3626-3634, 1992
A:Title: Solution structure of the fifth repeat of factor H: A second example of the complement factor H
A:Reference number: A44551; MUID:92232649
A:Contents: annotation: NMR structure determination, residues 264-292
R:Kristensen, T.; Wetsel, R.A.; Tack, B.F.
J. Immunol. 136, 3407-3411, 1986
A:Title: Structural analysis of human complement protein H: homology with C4b binding protein
A:Reference number: S10479; MUID:96169701
A:Accession: S10479
A:Molecule type: mRNA
A:Residues: 226-401, 'Y', 403-449 <KRI>
A:Cross-references: GB:M2383; NID:g180472; PID:AAAS2013.1; PID:g180473
C:Comment: Factor H has also been found bound to cell membranes in an unknown manner. H
C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver.
C:Genetics: <HF1>
A:Gene: GDB:HF1; HF
A:Cross-references: GDB:120041; OMIM:134370
A:Map position: 1q32-1q32
C:Genetics: <HF2>
A:Gene: GDB:HF2; HF
A:Cross-references: GDB:129095
A:Map position: 1q32-1q32
A:Note: the correspondence between the two loci and the sequences indicated is unclear;
C:Function:
A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also increased
he alternative complement pathway
A:Pathway: complement alternate pathway
C:Superfamily: complement factor H; complement factor H repeat homology
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-449/Product: complement factor H, short splice form #status experimental <MAT>
F:21-80/Domain: complement factor H repeat homology <FH01>
F:85-141/Domain: complement factor H repeat homology <FH02>
F:146-205/Domain: complement factor H repeat homology <FH03>
F:210-262/Domain: complement factor H repeat homology <FH04>
F:246-248/Region: cell attachment (R-G-D) motif
F:267-320/Domain: complement factor H repeat homology <FH05>
F:325-385/Domain: complement factor H repeat homology <FH06>
F:389-442/Domain: complement factor H repeat homology <FH07>
F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357
F:217/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 100.0%; Score 1497; DB 1; Length 449;
Best Local Similarity 100.0%; Pred. No. 7.9e-106;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDCNELPPRRNTEILTGSWSQDTYPEGTQAIYKCRPGYRSLGNVIMVCRKEWVALNPLR 60
Db 19 EDCNELPPRRNTEILTGSWSQDTYPEGTQAIYKCRPGYRSLGNVIMVCRKEWVALNPLR 78
QY 61 KCQRKPCGHPGDPFGFTTGTGNNVFEYGVKAVYTCNEGQYLLGEINRECDTGDGTNDI 120
Db 79 KCQRKPCGHPGDPFGFTTGTGNNVFEYGVKAVYTCNEGQYLLGEINRECDTGDGTNDI 138
QY 121 PICEVKKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVNCYKKTGDEMHCSDDGF 180
Db 139 PICEVKKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVNCYKKTGDEMHCSDDGF 198
QY 181 SKKPKCVEISKSPDVINGSPISQKIYKENERFQYKCMNGYEYSGRDAVCTESGWRP 240
Db 199 SKKPKCVEISKSPDVINGSPISQKIYKENERFQYKCMNGYEYSGRDAVCTESGWRP 258
QY 241 LPSCEEKSCDNPYPNGDYSPLRIK 265
Db 259 LPSCEEKSCDNPYPNGDYSPLRIK 283

RESULT 2
NBHUT
complement factor H precursor, long splice form [validated] - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 08-Dec-2000

C:Accession: S00254; A60238; A54726; A61565; A26505; I72654; S66298
R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.
Biochem. J. 249, 593-602, 1988
A:Title: The complete amino acid sequence of human complement factor H.
A:Reference number: S00254; MUID:88134059
A:Accession: S00254
A:Molecule type: mRNA
A:Residues: 1-1231 <RIP>
A:Cross-references: EMBL:Y00716; NID:g31964; PIDN:CAA68704.1; PID:g31965
A:Note: 402-Tyr was also found
A:Note: parts of this sequence, including the amino and carboxyl ends of the mature p
R:Estaller, C.; Schwaebble, W.; Dierich, M.; Weiss, E.H.
Eur. J. Immunol. 21, 799-802, 1991
A:Title: Human complement factor H: two factor H proteins are derived from alternativ
A:Reference number: A60238; MUID:91184292
A:Accession: A60238
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-56;1177-1231 <EST>
A:Note: only portions of this 4.3 kilobase mRNA were sequenced
R:Day, A.J.; Ripoche, J.; Lyons, A.; McIntosh, B.; Harris, T.J.R.; Sim, R.B.
Biochem. Rep. 7, 201-207, 1987
A:Title: Sequence analysis of a cDNA clone encoding the C-terminal end of human compl
A:Reference number: A54726; MUID:88025472
A:Accession: A54726
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-579-1231 <DAY>
A:Cross-references: GB:M17517; NID:g180497; PIDN:AAAS2016.1; PID:g180498
R:Ripoche, J.; Day, A.J.; Willis, A.C.; Belt, K.T.; Campbell, R.D.; Sim, R.B.
Biochem. Rep. 6, 65-72, 1986
A:Title: Partial characterization of human complement factor H by protein and cDNA se
A:Reference number: A61565; MUID:86188123
A:Accession: A61565
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1050-1057, 'T', 1059-1102 <RI2>
R:Sim, R.B.; DiScipio, R.G.
Biochem. J. 205, 285-293, 1982
A:Title: Purification and structural studies on the complement-system control protein
A:Reference number: A26505; MUID:83048213
A:Accession: A26505
A:Molecule type: protein
A:Residues: 19-20, 'Q', 22-23, 'V', 31-33, 'Q', 35 <SIM>
R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, F
Biochemistry 31, 3626-3634, 1992
A:Title: Solution structure of the fifth repeat of factor H: A second example of the
A:Reference number: A44551; MUID:92232649
A:Contents: annotation: NMR structure determination, residues 264-292
R:Norman, D.G.; Barlow, P.N.; Baron, M.; Day, A.J.; Sim, R.B.; Campbell, I.D.
J. Mol. Biol. 219, 717-725, 1991
A:Title: Three-dimensional structure of a complement control protein module in soluti
A:Reference number: A49224; MUID:91278097
A:Contents: annotation: NMR structure determination, residues 927-985
R:Estaller, C.; Kostiainen, V.; Schwaebble, W.; Dierich, M.P.; Weiss, E.H.
J. Immunol. 146, 3190-3196, 1991
A:Title: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a nc
A:Reference number: I56100; MUID:91201892
A:Accession: I72654
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1047-1231 <RES>
A:Cross-references: GB:M65294; NID:g183766; PIDN:AAA35948.1; PID:g183767
R:Carroll, J.A.; Bates, R.C.; Smith, A.I.; Tetoz, T.; Arellano, A.; Gordon, D.L.; Burn
Biochim. Biophys. Acta 1289, 305-311, 1996
A:Title: Factor H co-purifies with thrombospondin isolated from platelet secretate.
A:Reference number: S66298; MUID:96205365
A:Accession: S66298
A:Status: preliminary
A:Molecule type: protein
A:Residues: 411-419,574-578,580-582 <CAR>

C:Comment: Factor H has also been found bound to cell membranes in an unknown manner. H
C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver.
C:Genetics: <HF1>
A:Gene: GDB:HF1; HF
A:Cross-references: GDB:120041; OMIM:134370
A:Map position: lq32-lq32
C:Genetics: <HF2>
A:Gene: GDB:HF2; HF
A:Cross-references: GDB:129095
A:Map position: lq32-lq32
A:Note: the correspondence between the two loci and the sequences indicated is unclear;
C:Function:
A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also increases
he alternative complement pathway
A:Pathway: complement alternate pathway
C:Superfamily: complement factor H; Complement factor H repeat homology
C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-1229/Product: complement factor H #status experimental <MPT>
F:19-449/Product: complement factor H, short splice form #status experimental <MAT>
F:21-80/Domain: complement factor H repeat homology <FH01>
F:85-141/Domain: complement factor H repeat homology <FH02>
F:146-205/Domain: complement factor H repeat homology <FH03>
F:210-262/Domain: complement factor H repeat homology <FH04>
F:246-248/Region: cell attachment (R-G-D) motif
F:267-320/Domain: complement factor H repeat homology <FH05>
F:325-385/Domain: complement factor H repeat homology <FH06>
F:389-442/Domain: complement factor H repeat homology <FH07>
F:448-505/Domain: complement factor H repeat homology <FH08>
F:509-564/Domain: complement factor H repeat homology <FH09>
F:569-623/Domain: complement factor H repeat homology <FH10>
F:630-684/Domain: complement factor H repeat homology <FH11>
F:691-744/Domain: complement factor H repeat homology <FH12>
F:753-803/Domain: complement factor H repeat homology <FH13>
F:811-864/Domain: complement factor H repeat homology <FH14>
F:870-926/Domain: complement factor H repeat homology <FH15>
F:931-984/Domain: complement factor H repeat homology <FH16>
F:989-1043/Domain: complement factor H repeat homology <FH17>
F:1048-1102/Domain: complement factor H repeat homology <FH18>
F:1109-1163/Domain: complement factor H repeat homology <FH19>
F:1167-1228/Domain: complement factor H repeat homology <FH20>
F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357-
1-803,811-853,839-864,870-915,901-926,931-973,959-984,989-1032,1038-1043,1048-1091,1077-
F:217/Binding site: carbohydate (Asn) (covalent) #status absent
F:219,802,822,882,911/Binding site: carbohydate (Asn) (covalent) #status experimental
F:718,1029,1095/Binding site: carbohydate (Asn) (covalent) #status predicted

Query Match	100.0%;	Score 1497;	DB 1;	Length 1231;	
Best Local Similarity	100.0%;	Pred. No. 2.3e-105;			
Matches 265;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	EDCNELPPRRNTEILTGSWS	DQTYPGTQAIIYKCRPGYRSLGNVIMVCRKGEWVALNPLR	60	
Db					
	19	EDCNELPPRRNTEILTGSWS	DQTYPGTQAIIYKCRPGYRSLGNVIMVCRKGEWVALNPLR	78	
QY	61	KCOKRPGCHGDPDPFGTFTLT	GNVFEYGVKAVYTCNCGYQLLGEINYNRECDT	120	
Db					
	79	KCOKRPGCHGDPDPFGTFTLT	GNVFEYGVKAVYTCNCGYQLLGEINYNRECDT	138	
QY	121	PICEVVKCLPVTAPENKGI	VSSAMEPDRYHFQGA	VRFCNSGYKIEGDEEMHCSDDGFW	180
Db					
	139	PICEVVKCLPVTAPENKGI	VSSAMEPDRYHFQGA	VRFCNSGYKIEGDEEMHCSDDGFW	198
QY	181	SKEKPKCIVEISCKSPD	IVNGSPISQRIIYKENERFOYKCNMGY	EYYSERGDVCTESGWRP	240
Db					
	199	SKEKPKCIVEISCKSPD	IVNGSPISQRIIYKENERFOYKCNMGY	EYYSERGDVCTESGWRP	258
QY	241	LPSCEKSKCDNPYP	IPNGDYSPLRIK	265	
Db					
	259	LPSCEKSKCDNPYP	IPNGDYSPLRIK	283	

RESULT 3

NBMSH

Complement factor H precursor - mouse

N:Alternate names: protein beta-1-H

C:Species: Mus musculus (house mouse)

C:Date: 30-Sep-1987 #sequence.revision 30-Sep-1987 #text_change 22-Jun-1999

C:Accession: A26154; I49711; I49728

R:Kristensen, T.; Tack, B.F.

Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986

A:Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in length

A:Reference number: A26154; MUID:86233353

A:Accession: A26154

A:Molecule type: mRNA

A:Residues: 1-1234 <KRI>

A:Cross-references: GB:M12660; NID:g193724; PIDN:AAA37759.1; PID:g387101

R:Natsume-Sakai, S.; Nonaka, M.; Nonaka, Y.

J. Immunol. 144, 358-362, 1990

A:Title: Demonstration of an unusual allelic variation of mouse factor H by the complement

A:Reference number: I49711; MUID:90111033

A:Accession: I49711

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-18 <RES>

A:Cross-references: GB:M31979; NID:g193726; PIDN:AAA37762.1; PID:g193729

R:Munoz-Canoves, P.; Tack, B.F.; Vik, D.P.

Biochemistry 28, 9891-9897, 1989

A:Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma

A:Reference number: I49728; MUID:90148935

A:Accession: I49728

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-19 <RE2>

A:Cross-references: GB:J02891; NID:g193805; PIDN:AAA37795.1; PID:g553026

C:Comment: Two dominant alleles of factor H are present in mice.

C:Comment: Factor H functions as a cofactor in the inactivation of C3b by serine protease (C5 convertase) in the alternative complement pathway.

C:Genetics:

A:Map position: 1
C:Superfamily: complement factor H; complement factor H repeat homology
C:Keywords: complement alternate pathway; duplication; glycoprotein; plasma
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-1234/Product: complement factor H #status predicted <MPT>
F:21-80/Domain: complement factor H repeat homology <FH01>
F:85-141/Domain: complement factor H repeat homology <FH02>
F:146-205/Domain: complement factor H repeat homology <FH03>
F:210-262/Domain: complement factor H repeat homology <FH04>
F:246-248/Region: c5i1 attachment (R-G-D) motif
F:267-320/Domain: complement factor H repeat homology <FH05>
F:325-385/Domain: complement factor H repeat homology <FH06>
F:389-442/Domain: complement factor H repeat homology <FH07>
F:448-505/Domain: complement factor H repeat homology <FH08>
F:509-564/Domain: complement factor H repeat homology <FH09>
F:569-622/Domain: complement factor H repeat homology <FH10>
F:629-683/Domain: complement factor H repeat homology <FH11>
F:690-743/Domain: complement factor H repeat homology <FH12>
F:752-802/Domain: complement factor H repeat homology <FH13>
F:808-861/Domain: complement factor H repeat homology <FH14>
F:867-931/Domain: complement factor H repeat homology <FH15>
F:936-989/Domain: complement factor H repeat homology <FH16>
F:994-1048/Domain: complement factor H repeat homology <FH17>
F:1053-1107/Domain: complement factor H repeat homology <FH18>
F:1114-1168/Domain: complement factor H repeat homology <FH19>
F:1172-1233/Domain: complement factor H repeat homology <FH20>
F:121-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,
0-802,808-850,836-861,867-920,906-931,936-978,964-989,994-1037,1033-1048,1053-1096,10
F:676,721,773,801,1030,1061,1225/Binding site: carbohydrate (Asn) (covalent) #status

Query Match	68.4%;	Score 1024;	DB 1;	Length 1234;
Best Local Similarity	67.0%;	Pred. No. 1.3e-69;		

QY 1 EDCNELPPRRNTEILLTGSWSDQTYPEGTOAIYKCRPGVSLGNIMVCRKGEWALNPLR 60

Db 19 EDCKGPPPRENSEITLSSWSEQLYPECTQATYKRCPCGYRTLTGTVKCKKQWVASPSR 78
QY 61 KCKRCPCGHPGDPFGTFTLTGTVNVEYGVKAVYTCNEGQVLLGEINRECDTGDGTNDI 120
Db 79 ICRKKCGHPGDPFGFSRLAVGSQFEFGAKVYTCDDGYLLGEIDYRCGADGWINDI 138
QY 121 PICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRVCNSGYKIEGDEMHCSDDGFW 180
Db 139 PICEVVKCLPVTLENGRIVSGRAETDQYFFGVQVRFECNSGFKIEGHEHCSENGLW 198
QY 181 SKKPKCVETSCSPDVINGSPISQIKIYKENERFQYKCMNGYEYSERGDVACTESGWRP 240
Db 199 SNEKPRCVETLTPPRVINGDINVKPVYKENERYHYKCKHGYVPRGERGDVACTGSGWS 258
QY 241 LPSCEEKSCDNPYPNGDYSPLRI 264
Db 259 QPCEERKCSPPYLNGIYTPHRI 282

RESULT 4
S65551
factor H - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S65551
R:Soames, C.J.; Day, A.J.; Sim, R.B.
Biochem. J. 315, 523-531, 1996
A:Title: Prediction from sequence comparisons of residues of factor H involved in the in
A:Reference number: S65551, MUID:96202005
A:Accession: S65551
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-669 <SOA>
A:Cross-references: GB:X98697; NID:g1419423; PIDN:CAA67257.1; PID:g1419424
C:Superfamily: complement factor H; complement factor H repeat homology
F:55-114/Domain: complement factor H repeat homology <FHR1>
F:296-349/Domain: complement factor H repeat homology <FHR2>
F:355-412/Domain: complement factor H repeat homology <FHR3>
F:416-471/Domain: complement factor H repeat homology <FHR4>
F:476-530/Domain: complement factor H repeat homology <FHR5>
F:538-592/Domain: complement factor H repeat homology <FHR6>
F:599-651/Domain: complement factor H repeat homology <FHR6>
Query Match 46.7%; Score 699; DB 2; Length 669;
Best Local Similarity 63.4%; Pred. No. 2.6e-45;
Matches 118; Conservative 26; Mismatches 42; Indels 0; Gaps 0;

QY 76 GTFTLTGGNVFYGKAVYTCNEGQVLLGEINRECDTGDGTNDIPICEVVKCLPVTAPE 135
Db 3 GSPLAEGNOFEGYKAVYTCDEGYQVGMGNFRECDTNGWTNDIPICEVVKCLPVTPE 62
QY 136 NGKIVSSAMEPDREYHFGQAVRVCNSGYKIEGDEMHCSDDGFWSEKPKCVETSCSKSP 195
Db 63 NGKIFSDALEPDQYNYGVQVQFECNSGYMLDQPKIHCAGVSWAETPKCVETICKPP 122
QY 196 DVINGSPISQIKIYKENERFQYKCMNGYEYSERGDVACTESGWRPLPSCSEKSCDNPYIP 255
Db 123 VILNGQAVLPKATYKQNERVQYRCAGAFYGGQRGDTVCTKSGWTPAPTCTEITCDPPRIP 182
QY 256 NGDYSYP 261
Db 183 NGVYRP 188

RESULT 5
S46199
probable complement regulatory plasma protein SBI - barred sand bass
C:Species: Paralabrax nebulifer
C:Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 18-Jul-2001
C:Accession: S46199; S77894
R:Dahmen, A.; Kaidoh, T.; Zipfel, P.F.; Gigli, I.

Biochem. J. 301, 391-397, 1994
A:Title: Cloning and characterization of a cDNA representing a putative complement-re
A:Reference number: S46199; MUID:94318039
A:Accession: S46199
A:Molecule type: mRNA
A:Residues: 1-1053 <DAH1>
A:Cross-references: EMBL:L21703; NID:g639894; PIDN:AAA92556.1; PID:g639895
A:Experimental source: liver
A:Accession: S77894
A:Molecule type: protein
A:Residues: 526-532, 'X', 534-537; 809-817, 'X', 819-826 <DAH2>
C:Genetics:
A:Gene: SBI
C:Superfamily: complement factor H repeat homology
C:Keywords: glycoprotein
F:89-145/Domain: complement factor H repeat homology <FH01>
F:334-389/Domain: complement factor H repeat homology <FH02>
F:450-502/Domain: complement factor H repeat homology <FH03>
F:569-624/Domain: complement factor H repeat homology <FH03>
F:682-738/Domain: complement factor H repeat homology <FH04>
F:743-802/Domain: complement factor H repeat homology <FHR>
F:935-989/Domain: complement factor H repeat homology <FH05>
F:993-1052/Domain: complement factor H repeat homology <FH06>

Query Match 29.4%; Score 440; DB 2; Length 1053;
Best Local Similarity 36.2%; Pred. No. 1.6e-25;
Matches 89; Conservative 35; Mismatches 108; Indels 14; Gaps 6;
QY 21 DQTYPEGTQAIYKRCPCGYRSLGNVIMYCRKGWVALNPLRCKCRKPCGDPFGFTTL 80
Db 47 EASYPGGQVRVCGNCVGS--GFFKLVCEVGKQWETRG--AKCQPRSCGHPGDAQFADPHL 102
QY 81 TGCNVREYGVKAVYTCNEGQVLLGEINRECDTGDGTNDIPICEVVKCLPVTAPENGKIV 140
Db 103 AEGNDVFGSKVYTCQKGYQMVSRINRRCVAGEGMDGVVYVPCESQCCPLIHVDNNQVVI 162
QY 141 SSAMEPDREYHFGQAVRVCNSGYKI-EGDEEMHCSDDGFWSEKPKCVETSCSKSPDVIN 199
Db 163 GG----PEEATFGNVYRFSCKSRSEILDGSPELYCDERGDWSPVVKCAITCAIPPIEN 218
QY 200 GSPISQIKIYKENERFQYKCMNGYEYSERGDVACTESG----WRPLPSCSEKSCDNPYIP 255
Db 219 GNVPGAIREYKENDVLHYECDRAFKHIDR-PSTCIKQIKAEWSPTPLCESIKRLTIMD 277
QY 256 NGDYSYP 261
Db 278 GTRYEP 283

RESULT 6
T42921
complement control protein homolog ccph - ateline herpesvirus 3 (strain 73)
C:Species: ateline herpesvirus 3
A:Variety: Strain 73
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 04-Mar-2000
C:Accession: T42921
R:Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A:Description: Primary structure of the herpesvirus ateles genome.
A:Reference number: Z22274
A:Accession: T42921
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-360 <ALB>
A:Cross-references: EMBL:AF083424; PIDN:AAC95530.1
A:Experimental source: strain 73
C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom

Query Match 22.2%; Score 332.5; DB 2; Length 360;
Best Local Similarity 34.9%; Pred. No. 6.9e-18;
Matches 96; Conservative 32; Mismatches 106; Indels 41; Gaps 18;

Db 119 --LRNQVEIKTDLSEFGSIEFSCSGFLIGSTTSR-CEVDQDGVGWSHPLPQCEIVK 175
QY 129 LPVTAPENGKIVSSAMEPDREYHFGQAVRFVNSGKIEGDEEMHCSDD---GFWSEK 184
Db 176 KPPDRLNRHSGE---ENFAYGFSVYSCDPRESLILGHASICTVNEIGVWRPSP 231
QY 185 PKCVELSKSPDVINGSPISO-KIYKENERFOYKCNMGYEYSEKDAY--C-TESGWRP 240
Db 232 PTCEKITCRKPDVSHGEMVSGFGPIYNYKDTIVFKCKQKGVFL--RGSSVIHCADSKWNP 289
QY 241 L-PSCEEKSCDN-PYIPNGDY 259
Db 290 SPFACEPNSCINLPDIPHASW 310

RESULT 8
S53711
C4BP alpha chain precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S53711
R:de Frutos, P.G.; Dahlbaeck, B.
Biochim. Biophys. Acta 1261, 285-289, 1995
A:Title: cDNA structure of rabbit C4b-binding protein alpha-chain. Preserved sequence mo
A:Reference number: S53711; MUID:95226458
A:Accession: S53711
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-597 <DEF>
A:Cross-references: EMBL:235490
F:50-107/Domain: complement factor H repeat homology <FH1>
F:112-169/Domain: complement factor H repeat homology <FH2>
F:174-234/Domain: complement factor H repeat homology <FH3>
F:239-294/Domain: complement factor H repeat homology <FH4>
F:299-360/Domain: complement factor H repeat homology <FH5>
F:364-422/Domain: complement factor H repeat homology <FH6>
F:426-480/Domain: complement factor H repeat homology <FH7>
F:484-538/Domain: complement factor H repeat homology <FH8>

Query Match 20.2%; Score 302; DB 1; Length 597;
Best Local Similarity 32.0%; Pred. No. 2,4e-15;
Matches 87; Conservative 37; Mismatches 118; Indels 30; Gaps 16;

QY 2 DNEIPLPRNTIELIGSWSDQYPTGTOAIYKRCPCYSLG-NVIMVCR-KGEWALNPL 59
Db 49 DCGP-PPHLLFASSISELSENEYQTILKYTCRPGYTRNGLNPLITCKPRLW---SVD 104
QY 60 RKCKRCPCHGCDTPGFTLTGNNVEYGVKAVYTCNEGQYLLGEINRECDTD----G 115
Db 105 TFCVKKRCNPGDLPNGQVEK--TDFSGSQIEFSCSEGYILIGSTT-SHCDI0EKGV 161
QY 116 WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDREYH-FGQAVRFVNSGKIEGDEEMHC 174
Db 162 WSDPLPKCEIVKCEPPNINLKHNGG---NEDIHTGSSVYSCNPRFSLIGEAISIC 217
QY 175 S-----DDGFWSEKPKCVELSKSPDVINGSPISO-KIYKENERFOYKCNMGYEYSEK 229
Db 218 TVKNKTGVWSPSPVPCKEILICSPNPVPHGKIISGFGPIYNYKDSIMVTCIDGFLV--RG 275
QY 230 DAV--C-TESGWRPLPS-CEEKSC-DNPYIPN 256
Db 276 SSLIHCLEDSKWNPPPPVCSNCLGLNPNVPH 307

RESULT 9
WNVZSP
apolipoprotein H homolog precursor - vaccinia virus
N:Alternate names: 35k secretory protein; C3L protein; virokinase
C:Species: vaccinia virus
A:Note: host Homo sapiens (man)
C>Date: 31-Dec-1989 #sequence_revision 30-Jun-1990 #text_change 22-Jun-1999

C:Accession: A31005; B42504
R:Kotwal, G.J.; Moss, B.
Nature 335, 176-178, 1988
A:Title: Vaccinia virus encodes a secretory polypeptide structurally related to compl
A:Reference number: A31005; MUID:88318974
A:Accession: A31005
A:Molecule type: DNA
A:Residues: 1-263 <KOT>
A:Cross-references: GB:X13166; NID:g60690; PIDN:CAA31564.1; PID:g60691
A:Experimental source: strain WR
R:Goebel, S.J.; Johnson, G.P.; Perkins, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 517-563, 1990
A:Title: Appendix to "The complete DNA sequence of vaccinia virus".
A:Reference number: A42501
A:Accession: B42504
A:Molecule type: DNA
A:Residues: 1-263 <GOE>
A:Cross-references: GB:M35027; NID:g335317; PIDN:AAA47997.1; PID:g335345
A:Experimental source: strain Copenhagen
R:Goebel, S.J.; Johnson, G.P.; Perkins, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 247-266, 1990
A:Title: The complete DNA sequence of vaccinia virus.
A:Reference number: A42531; MUID:91021027
A:Contents: annotation; possible protein-coding frames
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom
C:Keywords: duplication; extracellular protein
F:1-19/Domain: signal sequence status predicted <SIG>
F:20-263/Product: C4b-binding protein homolog #status predicted <SIG>
F:21-81/Domain: complement factor H repeat homology <FH1>
F:86-143/Domain: complement factor H repeat homology <FH2>
F:148-201/Domain: complement factor H repeat homology <FH3>
F:206-261/Domain: complement factor H repeat homology <FH4>

Query Match 20.1%; Score 301.5; DB 1; Length 263;
Best Local Similarity 33.3%; Pred. No. 1.1e-15;
Matches 78; Conservative 29; Mismatches 106; Indels 21; Gaps 12;

QY 20 SDQTYPEGTQAIYKCRPGYR--SLGNVIMVCRKGEWALNPLKCKRCPCHGCDTPFGT 77
Db 40 ANANTNIGDTIELYCLPGYKRMGMPIYAKCTGTGTLFN---QCIRKRCSPRDIENG 96
QY 78 FTLTGNNVEYGVKAVYTCNEGQYLLGEI-NYRECDTDG---WTNDIPICEVVKCLPVT 133
Db 97 LDIGG--VDFGSSITYSCNSGYHLIGESKVCYELGSGTGMWNPAPICESVKCQSPS 153
QY 134 PENGKIVSSAMEPDREYHFGQAVRFVNSGKIEGDEEMHCSDDGFWSEKPKCVELSK 193
Db 154 ISNGR--HNGYED--FYTDGSVVYTCNSGYSLIGNSGVLCS-GGEWS-DPPTCQIVKCP 207
QY 194 SPDVINGSPISO-KIYKENERFOYKCNMGYEYSEKDAYCTESG-WRP-LPSC 244
Db 208 HPTISNGYLSGFKRSYSDNWDPRKYGLSGSSSTCSGNTWKPELPRC 261

RESULT 10
A45900
complement C3b receptor type 2 long form precursor - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 08-Oct-1999
C:Accession: A45900; I48306
R:Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Weis, J.H.
J. Immunol. 144, 3581-3591, 1990
A:Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2
A:Reference number: A45900; MUID:90229754
A:Accession: A45900
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-676 <KUR>
A:Cross-references: GB:M36470
A:Experimental source: clone 31-1
R:Kim, Y.U.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.M.; Holers,

A:Residues: 1-683,'X',685-894,'A',896-1000,1451-1471,'X',1473-2489 <VIK2>
A:Cross-references: GB:L17418; NID:g306678; PIDN:AAB60694.1; PID:g306680
R:Wong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.;
J. Exp. Med. 169, 847-863, 1989
A:Title: Structure of the human Crl gene. Molecular basis of the structural and quant
A:Reference number: A47602; MUID:89176859
A:Accession: A47602
A:Molecule type: DNA
A:Residues: 1-41 <WON>
A:Cross-references: GB:X14893
R:Hourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.
J. Exp. Med. 168, 1255-1270, 1988
A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b r
type 1.
A:Reference number: S03291; MUID:89010527
A:Accession: S03291
A:Molecule type: mRNA
A:Residues: 26-584 <HOU>
A:Cross-references: EMBL:X14362; NID:g30197; PIDN:CAA32541.1; PID:g736240
A:Experimental source: Clone Crl-4
R:Klickstein, L.B.; Bartow, T.J.; Milletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T.
J. Exp. Med. 168, 1699-1717, 1988
A:Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4b r
A:Reference number: S03843; MUID:89035992
A:Accession: S03843
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-894,'A',896-1000,1451-2064,'I',2066-2276,'P',2278-2299,'H',2301-2325,'T',232
A:Cross-references: EMBL:Y00816; NID:g30185; PIDN:CAA68755.1; PID:g30186
R:Klickstein, L.B.; Wong, W.W.; Smith, J.A.; Weis, J.H.; Wilson, J.G.; Fearon, D.T.
J. Exp. Med. 165, 1095-1112, 1987
A:Title: Human C3b/C4b receptor (CRL). Demonstration of long homologous repeating dom
A:Reference number: A28507; MUID:87168191
A:Accession: A28507
A:Molecule type: mRNA
A:Residues: 953-1221,'FV',1224-2064,'I',2066-2276,'P',2278-2299,'H',2301-2325,'T',232
A:Cross-references: GB:X05309; NID:g30136; PIDN:CAA28933.1; PID:g809019
R:Wong, W.W.; Klickstein, L.B.; Smith, J.A.; Weis, J.H.; Fearon, D.T.
Proc. Natl. Acad. Sci. U.S.A. 82, 7711-7715, 1985
A:Title: Identification of a partial cDNA clone for the human receptor for complement
A:Reference number: A94073; MUID:86067975
A:Accession: A24748
A:Molecule type: mRNA
A:Residues: 311-333;729-745;831-845 <W03>
A:Cross-references: GB:M11569; NID:g180991; PIDN:AAA52297.1; PID:g180995; GB:M11617;
C:Genetics:
A:Gene: GDB:CRL; CD35
A:Cross-references: GDB:l119800; OMIM:l20620
A:Map position: lq32-lq32
A:Introns: 41/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2; 672/2
; 1484/2; 1531/1; 1646/1; 1705/1; 1739/2; 1768/1; 1844/1; 1904/1; 1937/2; 1966/1; 209
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C:Keywords: duplication; glycoprotein; transmembrane protein
F:1-41/Domain: signal sequence #status predicted <SIG>
F:42-2489/Product: complement C3b/C4b receptor, membrane-bound form #status predicted
F:42-584/Product: complement C3b/C4b receptor, secreted form #status predicted <MATS>
F:43-99/Domain: complement factor H repeat homology <FH01>
F:104-161/Domain: complement factor H repeat homology <FH02>
F:166-232/Domain: complement factor H repeat homology <FH03>
F:238-293/Domain: complement factor H repeat homology <FH04>
F:297-353/Domain: complement factor H repeat homology <FH05>
F:358-416/Domain: complement factor H repeat homology <FH06>
F:421-487/Domain: complement factor H repeat homology <FH07>
F:493-549/Domain: complement factor H repeat homology <FH08>
F:554-611/Domain: complement factor H repeat homology <FH09>
F:616-682/Domain: complement factor H repeat homology <FH10>
F:688-743/Domain: complement factor H repeat homology <FH11>
F:747-803/Domain: complement factor H repeat homology <FH12>
F:808-866/Domain: complement factor H repeat homology <FH13>
F:871-937/Domain: complement factor H repeat homology <FH14>
F:943-999/Domain: complement factor H repeat homology <FH15>
F:1004-1061/Domain: complement factor H repeat homology <FH16>
F:1066-1132/Domain: complement factor H repeat homology <FH17>

F:37-488/Product: complement regulatory protein, 512 antigen #status predicted <MAT>
F:38-94/Domain: complement factor H repeat homology <FH1>
F:99-156/Domain: complement factor H repeat homology <FH2>
F:161-227/Domain: complement factor H repeat homology <FH3>
F:233-288/Domain: complement factor H repeat homology <FH4>
F:294-352/Domain: complement factor H repeat homology <FH5>
F:357-413/Domain: complement factor H repeat homology <FH6>
F:421-444/Domain: transmembrane #status predicted <TM>
F:247,331,346,450,482,483/Binding site: carbohydrate (Asn) #status predicted

Query Match 19.3%; Score 289.5; DB 2; Length 497;
Best Local Similarity 30.1%; Pred. No. 1.7e-14;
Matches 78; Conservative 42; Mismatches 108; Indels 31; Gaps 15;

QY 21 SDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPLRKCKQKRCPCGHPGDTFFGT 80
Db 55 ESTFPVGTSLKYECPGKRYIKRQFSITCEVNSVWT--SPQDVGIRKQCEPLDPQNGIVHV 112
QY 81 TCGNVEYGVKAVYTCNEGYYQLLGEINRECDTD---GWTNDIPICEVVKC-LPVTAPEN 136
Db 113 --NTDIRFGSSITYTCNEGYYRLIGSSAMCIIISDQSVAMDAEAPICESIPCEIIPPSP-N 169
QY 137 GKTIVSSAMEPDRE-YHFGQAVRVCVNSG-----YKIEGDEMHCS-----DDGFWSEKXP 185
Db 170 GDFFS-----PNREDFHYGVVITVQCTMDARGKKLFNLVGEPSIHCYSIDGQGVGWSGPPP 225
QY 186 KCVEIS-CKSPDVINGSPISQ-KIIYKENERFOYKCNMGYSEYSERGDAVCTE-SGWRP-L 241
Db 226 QCIELNKCTPPHVENAVIVSKNLSLRLDMVEFRCDGPFMMKGDSSVYCRSLNRWEPQL 285
QY 242 PSC-EKESCDN--PYTPNG 257
Db 286 PSCFKVKSCGAFIGELPNG 304

RESULT 14
C36838
complement control protein homolog - variola virus (strain India-1967)
N:Alternate names: D12L protein
C:Species: variola virus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 23-Mar-2001
R:Blinov, V.M.
submitted to GenBank, November 1992
A:Reference number: A36859
A:Accession: C36838
A:Molecule type: DNA
A:Residues: 1-263 <BLI>
A:Cross-references: GB: X69198; NID: q456758; PIDN: CAA48953.1; PID: q297195
C:Superfamily: herpesvirus complement control protein; complement factor H repeat homology <FH1>
F:21-81/Domain: complement factor H repeat homology <FH1>
F:86-143/Domain: complement factor H repeat homology <FH2>
F:148-201/Domain: complement factor H repeat homology <FH3>
F:206-261/Domain: complement factor H repeat homology <FH4>

Query Match 19.0%; Score 284.5; DB 1; Length 263;
Best Local Similarity 32.1%; Pred. No. 2.1e-14;
Matches 75; Conservative 29; Mismatches 109; Indels 21; Gaps 11;

QY 20 SDQTYPEGTQAIYKCRPGYR--SLGNVIMVCRKGEWVALNPLRKCKQKRCPCGHPGDTFFGT 77
Db 40 ANANYNIGDTIEYLCPLPGYRKQKMPIYAKCTGTGTMFLN---QCIRKRCPSPRDIDNGH 96
QY 78 FTLTGNNVEYGVKAVYTCNEGYYQLLGEI-NYRECDTDG---WTNDIPICEVVKCLPVT 133
Db 97 LDIGG---VDFGSSITYSCNSGYLLIGEYKSYCKLGSTGSMVWNPAPICESYKCOLPPS 153
QY 134 PENGKIVSSAMEPDREYHFGQAVRVCVNSGYKIEGDEMHCSDDGFWSEKPKCQVEISCK 193
Db 154 ISNGRHNGY----NDFYTDGSVYVTCNSGYSLIGNSGVLCS-GGEWS-NPPTCQIVKCP 207

Search completed: August 30, 2002, 06:18:08
Job time: 802 sec

QY 194 SPDVINGSPISQ-KIIYKENERFOYKCNMGYSEYSERGDAVCTESG-WRP-LPSC 244
Db 208 HPTILNGLSSGFKRSYNDNVDFTCYGYKLSGSSSSSTCSPGNTWOPELPKC 261

RESULT 15
T28450
hypothetical protein D15L - variola major virus
C:Species: variola major virus
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
R:Assung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Uterback, T.R.; Knight, J.C.; Au
Nature 366, 748-751, 1993
A:Title: Potential virulence determinants in terminal regions of variola smallpox vir
A:Reference number: 220488; MUID: 94088747
A:Accession: T28450
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-263 <MAS>
A:Cross-references: EMBL: L22579; NID: g623595; PIDN: AAA60760.1; PID: g438930
A:Experimental source: strain "Bangladesh-1975"
C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom

Query Match 19.0%; Score 284.5; DB 2; Length 263;
Best Local Similarity 32.1%; Pred. No. 2.1e-14;
Matches 75; Conservative 29; Mismatches 109; Indels 21; Gaps 11;

QY 20 SDQTYPEGTQAIYKCRPGYR--SLGNVIMVCRKGEWVALNPLRKCKQKRCPCGHPGDTFFGT 77
Db 40 ANANYNIGDTIEYLCPLPGYRKQKMPIYAKCTGTGTMFLN---QCIRKRCPSPRDIDNGH 96
QY 78 FTLTGNNVEYGVKAVYTCNEGYYQLLGEI-NYRECDTDG---WTNDIPICEVVKCLPVT 133
Db 97 LDIGG---VDFGSSITYSCNSGYLLIGEYKSYCKLGSTGSMVWNPAPICESYKCOLPPS 153
QY 134 PENGKIVSSAMEPDREYHFGQAVRVCVNSGYKIEGDEMHCSDDGFWSEKPKCQVEISCK 193
Db 154 ISNGRHNGY----NDFYTDGSVYVTCNSGYSLIGNSGVLCS-GGEWS-NPPTCQIVKCP 207

C 194 SPDVINGSPISQ-KIIYKENERFOYKCNMGYSEYSERGDAVCTESG-WRP-LPSC 244
Db 208 HPTILNGLSSGFKRSYNDNVDFTCYGYKLSGSSSSSTCSPGNTWOPELPKC 261

Search completed: August 30, 2002, 06:18:08
Job time: 802 sec

GenCore version 4.5
 Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2002, 06:17:06 ; Search time 28.81 Seconds
 (without alignments)
 356.150 Million cell updates/sec

Title: US-09-316-163-10
 Perfect score: 1497
 Sequence: 1 EDCNELPRRTEILGTSWS.....EKSCDNPYPNGDYSPLRIK 265

Scoring table:
 BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1497	100.0	1231	1 CFAH_HUMAN	P08603 homo sapien
2	1024	68.4	1234	1 CFAH_MOUSE	P06909 mus musculus
3	723.5	48.3	685	1 CFAH_BOVIN	Q28085 bos taurus
4	303	20.2	597	1 C4BP_HUMAN	P04003 homo sapien
5	301.5	20.1	263	1 VCP_VACCV	P10998 vaccinia vi
6	298.5	19.9	2039	1 CRI_HUMAN	P17927 homo sapien
7	276.5	18.5	558	1 C4BP_RAT	Q63514 rattus norv
8	275	18.4	469	1 C4BP_MOUSE	P08607 mus musculus
9	274	18.3	377	1 MCP_HUMAN	P15329 homo sapien
10	272	18.2	610	1 LEM2_HUMAN	P16581 homo sapien
11	266	17.8	484	1 LEM2_PIG	P98110 sus scrofa
12	265.5	17.7	610	1 C4BP_BOVIN	Q28065 bos taurus
13	264.5	17.7	360	1 C4BP_HSVSA	Q01016 herpesvirus
14	262	17.5	381	1 DAF_HUMAN	P08174 homo sapien
15	260.5	17.4	768	1 LEM3_MOUSE	Q01102 mus musculus
16	260	17.4	612	1 LEM2_MOUSE	Q00690 mus musculus
17	257.5	17.2	1025	1 CR2_MOUSE	P19070 mus musculus
18	254	17.0	768	1 LEM3_RAT	P98106 rattus norv
19	254	17.0	1033	1 CR2_HUMAN	P20023 homo sapien
20	253	16.9	345	1 APOH_MOUSE	Q01339 mus musculus
21	252	16.8	340	1 DAF_PONPY	P49457 pongo pygma
22	252	16.8	661	1 LEM3_HUMAN	P05160 homo sapien
23	250.5	16.7	830	1 LEM3_HUMAN	P16109 homo sapien
24	249.5	16.7	769	1 LEM3_SHEEP	P98109 ovis aries
25	246.5	16.5	507	1 DAF_CAVPO	Q60401 cavia porce
26	241	16.1	549	1 LEM2_RAT	P98105 rattus norv
27	241	16.1	611	1 LEM2_CANFA	P33730 canis fami
28	240	16.0	345	1 APOH_BOVIN	P17690 bos taurus
29	239.5	16.0	551	1 LEM2_RABIT	P27113 oryctolagus
30	234	15.6	668	1 F13B_MOUSE	Q07968 mus musculus
31	233.5	15.6	958	1 HIG_DROME	Q09101 drosophila
32	231.5	15.5	390	1 DAF1_MOUSE	Q61475 mus musculus
33	231	15.4	345	1 APOH_HUMAN	P02749 homo sapien

34 222.5 14.9 407 1 DAF2_MOUSE Q61476 mus musculus

35 222.5 14.9 485 1 LEM2_BOVIN P98107 bos taurus

36 222 14.8 345 1 APOH_CANFA P33703 canis fami

37 216.5 14.5 646 1 LEM3_BOVIN P42201 bos taurus

38 210 14.0 372 1 LEM1_MOUSE P18337 mus musculus

39 203.5 13.6 372 1 LEM1_HUMAN P14151 homo sapien

40 203.5 13.6 372 1 LEM1_PANTR Q95237 pan troglod

41 202.5 13.5 372 1 LEM1_PONPY Q95235 pongo pygma

42 200 13.4 372 1 LEM1_RAT P30836 rattus norv

43 194 13.0 372 1 LEM1_MACMU Q95198 macaca mula

44 194 13.0 372 1 LEM1_PAPHA Q28768 papio hamad

45 192.5 12.9 270 1 FHR2_HUMAN P36980 homo sapien

ALIGNMENTS

RESULT 1

CPAH_HUMAN

ID CFAH_HUMAN STANDARD; PRT; 1231 AA.

AC P08603; Q14570; P78435; Q9NU86;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Complement factor H precursor (H factor 1).

GN HFI OR HF OR CFH.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID:9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT Y-402.

RC TISSUE=Liver;

RX MEDLINE=88134059; PubMed=2963625;

RA Ripocher J., Day A.J., Harris T.J.R., Sim R.B.;

RT "The complete amino acid sequence of human complement factor H.;"

RL Biochem. J. 249:593-602(1988).

RN [2]

RP SEQUENCE OF 53-445 FROM N.A.

RX MEDLINE=87054207; PubMed=2946589;

RA Schulz T.F., Schwaebler W., Stanley K.K., Weiss E., Dierich M.P.;

RT "Human complement factor H: isolation of cDNA clones and partial cDNA sequence of the 38-kDa tryptic fragment containing the binding site for C3b.;"

RL Eur. J. Immunol. 16:1351-1355(1986).

RN [3]

RP SEQUENCE OF 226-449 FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE=86169701; PubMed=2937845;

RA Kristensen T., Wetzel R.A., Tack B.F.;

RT "Structural analysis of human complement protein H: homology with C4b binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.;"

RL J. Immunol. 136:3407-3411(1986).

RN [4]

RP SEQUENCE FROM N.A. (ISOFORM 2).

RA Bird C.;

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE OF 1047-1231 FROM N.A.

RX MEDLINE=91201892; PubMed=1826708;

RA Estaller C., Koistinen V., Schwaebler W., Dierich M.P., Weiss E.H.;

RT "Cloning of the 1.4-kb mRNA species of human complement factor H reveals a novel member of the short consensus repeat family related to the carboxy terminal of the classical 150-kDa molecule.;"

RL J. Immunol. 146:3190-3196(1991).

RN [6]

RP SEQUENCE OF 19-35.

RX MEDLINE=83048213; PubMed=6215918;

RA Sim R.B., Discipio R.G.;

RT "Purification and structural studies on the complement-system control protein beta 1H (Factor H).;"

RL Biochem. J. 205:285-293(1982).

RN [7]

RP SEQUENCE OF 1-19 FROM N.A.


```
Query Match 100.0%; Score 1497; DB 1; Length 1231;
Best Local Similarity 100.0%; Pred. No. 6.5e-117;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDCNELPPRTELTGSDQTYPECTQAIYKRCGYRSLGNVIMVCRKGWVALNPLR 60
Db 19 EDCNELPPRTELTGSDQTYPECTQAIYKRCGYRSLGNVIMVCRKGWVALNPLR 78
QY 61 KCQKRCGHPGDPFGFTLTGNGVEYGVKAVYTCNEGYYQLLGEINRYRCDDTGWMTNDI 120
Db 79 KCQKRCGHPGDPFGFTLTGNGVEYGVKAVYTCNEGYYQLLGEINRYRCDDTGWMTNDI 138
QY 121 PICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVRFVNCNGYKKEGDEMHCSDDGF 180
Db 139 PICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVRFVNCNGYKKEGDEMHCSDDGF 198
QY 181 SKKPKCIVEISCKSPDVINGSPISQKIYKENERFOYKCNMGVEYRSGDVAVTCESGWRP 240
Db 199 SKKPKCIVEISCKSPDVINGSPISQKIYKENERFOYKCNMGVEYRSGDVAVTCESGWRP 258
QY 241 LPSCEEKSCDNPYPNGDYSPLRIK 265
Db 259 LPSCEEKSCDNPYPNGDYSPLRIK 283

RESULT 2
CFAH_MOUSE STANDARD; PRT: 1234 AA.
AC P06909;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Complement factor H precursor (protein beta-1-H).
GN HF1 OR CFH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86233353; PubMed=2940596;
RA Kristensen T., Tack B.F.;
RT "Murine protein H is comprised of 20 repeating units, 61 amino acids
in length.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:3963-3967(1986).
RN [2]
RP SEQUENCE OF 1-19 FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=90148935; PubMed=2533512;
RA Munoz-Ganoves P., Tack B.F., Vik D.P.;
RT "Analysis of complement factor H mRNA expression: dexamethasone and
IFN-gamma increase the level of H in L cells.";
RL Biochemistry 28:9891-9897(1989).
RN [3]
RP SEQUENCE OF 1-18 FROM N.A.
RX MEDLINE=90111033; PubMed=2136885;
RA Natsume-Sakai S., Nonaka M., Nonaka M., Harada Y.N., Shreffler D.C.,
RA Moriwaki K.;
RT "demonstration of an unusual allelic variation of mouse factor H by
the complete cDNA sequence of the H.2 allotype.";
RL J. Immunol. 144:358-362(1990).
CC -1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF
C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE
C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3b)NBB COMPLEX (C5
CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.
CC -1- POLYMORPHISM: TWO CODOMINANT ALLELES OF FACTOR H ARE PRESENT IN
MICE.
CC -1- SIMILARITY: CONTAINS 20 SUSHI (SCR) DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M12660; AAA37759.1; -
CC EMBL: J02891; AAA37795.1; -
CC EMBL: M31979; AAA37762.1; -
CC PIR: A26154; NBM5H.
CC HSP: P08603; 1HF1.
CC MGI: 88385; Cfh.
CC InterPro: IPR000436; Sushi_SCR_CCP.
CC Pfam: P00084; sushi; 20.
CC SMART: SM0032; CCP; 20.
CC Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;
KW Signal.
FT SIGNAL 1 18 BY SIMILARITY.
FT CHAIN 19 1234 COMPLEMENT FACTOR H.
FT DOMAIN 20 81 SUSHI 1.
FT DOMAIN 84 142 SUSHI 2.
FT DOMAIN 145 206 SUSHI 3.
FT DOMAIN 209 263 SUSHI 4.
FT DOMAIN 266 321 SUSHI 5.
FT DOMAIN 324 386 SUSHI 6.
FT DOMAIN 388 443 SUSHI 7.
FT DOMAIN 447 506 SUSHI 8.
FT DOMAIN 508 565 SUSHI 9.
FT DOMAIN 568 623 SUSHI 10.
FT DOMAIN 628 684 SUSHI 11.
FT DOMAIN 689 744 SUSHI 12.
FT DOMAIN 751 803 SUSHI 13.
FT DOMAIN 807 862 SUSHI 14.
FT DOMAIN 866 932 SUSHI 15.
FT DOMAIN 935 990 SUSHI 16.
FT DOMAIN 993 1049 SUSHI 17.
FT DOMAIN 1052 1108 SUSHI 18.
FT DOMAIN 1113 1169 SUSHI 19.
FT DOMAIN 1171 1234 SUSHI 20.
FT DISULFID 21 66 BY SIMILARITY.
FT DISULFID 52 80 BY SIMILARITY.
FT DISULFID 85 129 BY SIMILARITY.
FT DISULFID 114 141 BY SIMILARITY.
FT DISULFID 146 192 BY SIMILARITY.
FT DISULFID 178 205 BY SIMILARITY.
FT DISULFID 210 251 BY SIMILARITY.
FT DISULFID 237 262 BY SIMILARITY.
FT DISULFID 267 309 BY SIMILARITY.
FT DISULFID 294 320 BY SIMILARITY.
FT DISULFID 325 374 BY SIMILARITY.
FT DISULFID 357 385 BY SIMILARITY.
FT DISULFID 389 431 BY SIMILARITY.
FT DISULFID 416 442 BY SIMILARITY.
FT DISULFID 448 494 BY SIMILARITY.
FT DISULFID 477 505 BY SIMILARITY.
FT DISULFID 509 553 BY SIMILARITY.
FT DISULFID 536 564 BY SIMILARITY.
FT DISULFID 569 610 BY SIMILARITY.
FT DISULFID 597 622 BY SIMILARITY.
FT DISULFID 629 672 BY SIMILARITY.
FT DISULFID 658 683 BY SIMILARITY.
FT DISULFID 690 732 BY SIMILARITY.
FT DISULFID 718 743 BY SIMILARITY.
FT DISULFID 752 791 BY SIMILARITY.
FT DISULFID 780 802 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 861 BY SIMILARITY.
FT DISULFID 867 920 BY SIMILARITY.
FT DISULFID 906 931 BY SIMILARITY.
FT DISULFID 936 978 BY SIMILARITY.
FT DISULFID 964 989 BY SIMILARITY.
FT DISULFID 994 1037 BY SIMILARITY.
FT DISULFID 1023 1048 BY SIMILARITY.
FT DISULFID 1033 1096 BY SIMILARITY.
```

```
FT DISULFID 1082 1107 BY SIMILARITY.
FT DISULFID 1114 1157 BY SIMILARITY.
FT DISULFID 1143 1168 BY SIMILARITY.
FT DISULFID 1172 1223 BY SIMILARITY.
FT DISULFID 1206 1233 BY SIMILARITY.
FT CARBOHYD 676 676 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 773 773 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1061 1061 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1225 1225 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1234 AA; 139082 MW; C5AC02F341B957F7 CRC64;

Query Match 68.4%; Score 1024; DB 1; Length 1234;
Best Local Similarity 67.0%; Pred. No. 1.5e-77;
Matches 177; Conservative 30; Mismatches 57; Indels 0; Gaps 0;

QY 1 EDCNELPPRNTTEILTGSWSDDQTYPEGTQAIYKCRGYRSLGNVIMVCRKGEWVALNPLR 60
Db 19 EDCCKGPPPRENSILSGNSEQLYPEGTQATYKCRGYRTLTGTVKCKNGKVASNPSR 78

QY 61 KCQRKCGHDPGDPFGTFLTGNNVEYGVKAVYTCNEGQVLLGEINRYRCDTGTNDI 120
Db 79 ICRKKPCGHPGDPFGSFRFVAVSQFEFGAKVYTCDDGYQLLGEIDYRCEGADGWINDI 138

QY 121 PICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVRVCNSGYKIEGDEMHCSDDGF 180
Db 139 PICEVVKCLPVTLENGRIVSGAETDQYFGQVVRFCNSGFKIEGHEIHCSENGLW 198

QY 181 SKEPKKVEISCKSPDVINGSPIQKIIYKENERFYKCNMGYEYSGRDAVCTESGWRP 240
Db 199 SNEKPRCVELCTPPRVENGSGINVRKPVYKENERHYKCKHGYPVPRGSDAVCTGSWS 258

QY 241 LPSCEKSCDNPIPNQDYSPLRI 264
Db 259 QPFCEKRCSPPIVILNGIYTPHRI 282

RESULT 3
ID CFAH_BOVIN STANDARD; PRT; 685 AA.
AC Q28085;
OC 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Complement factor H (H factor 1) (Fragments).
GN HFI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE OF 17-685 FROM N.A., AND SEQUENCE OF 1-16.
RC TISSUE=Liver;
RX MEDLINE=96202005; PubMed=8615824;
RA Soames C.J., Day A.J., Sim R.B.;
RT "Prediction from sequence comparisons of residues of factor H involved
RT in the interaction with complement component C3b.";
RL Biochem. J. 315:523-531(1996).
CC -!- FUNCTION: Factor H functions as a cofactor in the inactivation of
CC C3b by factor I and also increases the rate of dissociation of the
CC C3bb complex (C3 convertase) and the (C3b)NBB complex (C5
CC convertase) in the alternative complement pathway (By similarity).
CC -!- SIMILARITY: CONTAINS AT LEAST 13 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X98697; CAA67257.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 11.
DR SMART; SM00032; CCP; 11.
KW Complement alternate pathway; Plasma; Repeat; Sushi.
FT NON_TER 1
FT NON_CONS 16 17
FT DOMAIN <17 67 SUSHI 2.
FT DOMAIN 70 131 SUSHI 3.
FT DOMAIN 134 188 SUSHI 4.
FT DOMAIN 191 246 SUSHI 5.
FT DOMAIN 249 309 SUSHI 6.
FT DOMAIN 311 366 SUSHI 7.
FT DOMAIN 370 429 SUSHI 8.
FT DOMAIN 431 488 SUSHI 9.
FT DOMAIN 491 547 SUSHI 10.
FT DOMAIN 553 609 SUSHI 11.
FT DOMAIN 614 668 SUSHI 12.
FT DOMAIN 675 >685 SUSHI 13.
FT DISULFID 39 66 BY SIMILARITY.
FT DISULFID 71 117 BY SIMILARITY.
FT DISULFID 103 130 BY SIMILARITY.
FT DISULFID 135 176 BY SIMILARITY.
FT DISULFID 162 187 BY SIMILARITY.
FT DISULFID 192 234 BY SIMILARITY.
FT DISULFID 219 245 BY SIMILARITY.
FT DISULFID 250 297 BY SIMILARITY.
FT DISULFID 302 308 BY SIMILARITY.
FT DISULFID 312 354 BY SIMILARITY.
FT DISULFID 339 365 BY SIMILARITY.
FT DISULFID 371 417 BY SIMILARITY.
FT DISULFID 400 428 BY SIMILARITY.
FT DISULFID 432 476 BY SIMILARITY.
FT DISULFID 459 487 BY SIMILARITY.
FT DISULFID 492 534 BY SIMILARITY.
FT DISULFID 520 546 BY SIMILARITY.
FT DISULFID 554 597 BY SIMILARITY.
FT DISULFID 583 608 BY SIMILARITY.
FT DISULFID 615 656 BY SIMILARITY.
FT DISULFID 642 667 BY SIMILARITY.
FT NON_TER 685 685
SQ SEQUENCE 685 AA; 77536 MW; 69FC9DC8D530E872 CRC64;

Query Match 48.3%; Score 723.5; DB 1; Length 685;
Best Local Similarity 49.8%; Pred. No. 7.7e-53;
Matches 130; Conservative 28; Mismatches 46; Indels 57; Gaps 1;

QY 1 EDCNELPPRNTTEILTGSWSDDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPLR 60
Db 1 EDCCKGPPPRKTEILSVS----- 18

QY 61 KCQRKCGHDPGDPFGTFLTGNNVEYGVKAVYTCNEGQVLLGEINRYRCDTGTNDI 120
Db 19 -----GSPHLAEGNQFEYGAQVYTCDEGYQMGEMNFRECDTGTNDI 63

QY 121 PICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVRVCNSGYKIEGDEMHCSDDGF 180
Db 64 PICEVVKCLPVTPEPKGKIFSDALEPDQYTYGVQVQFECNSGYMLDGPQKHCSAGGV 123

QY 181 SKEPKKVEISCKSPDVINGSPIQKIIYKENERFYKCNMGYEYSGRDAVCTESGWRP 240
Db 124 SAETPRKVEIFCKPPVILNGQAVLPKATYKQNERVOYRCAAGFEYQGRGDTVCTKSGWTP 183

QY 241 LPSCEKSCDNPIPNQDYSPLRI 261
Db 184 APTCIEITCDPPRIPNGVYRP 204
```

RESULT 4
C4BP_HUMAN
ID C4BP_HUMAN STANDARD; PRT; 597 AA.
AC P04003;
DT 23-OCT-1986 (Rel. 02, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C4b-binding protein alpha chain precursor (C4bp) (Proline-rich
DE protein) (PRP).
DE C4BPA OR C4BP.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=90073699; PubMed=2590215;
RA Matsuguchi T., Okamura S., Aso T., Sata T., Niho Y.;
RT "Molecular cloning of the cDNA coding for proline-rich protein (PRP):
RL Identity of PRP as C4b-binding protein.";
RL Biochem. Biophys. Res. Commun. 165:138-144(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=91113199; PubMed=1989602;
RA Aso T., Okamura S., Matsuguchi T., Sakamoto N., Sata T., Niho Y.;
RT "Genomic organization of the alpha chain of the human C4b-binding
RL protein gene.";
RL Biochem. Biophys. Res. Commun. 174:222-227(1991).
RN [3]
RP SEQUENCE OF 9-81 FROM N.A.
RC MEDLINE=98242821; PubMed=3378624;
RA Lintin S.J., Lewin A.R., Reid K.B.M.;
RT "Derivation of the sequence of the signal peptide in human
RL C4b-binding protein and interspecies cross-hybridisation of the C4bp
RL cDNA sequence.";
RL FEBS Lett. 232:328-332(1988).
RN [4]
RP SEQUENCE OF 203-288 FROM N.A.
RC MEDLINE=86301119; PubMed=3017751;
RA Lintin S.J., Reid K.B.M.;
RT "Studies on the structure of the human C4b-binding protein gene.";
RL FEBS Lett. 204:77-81(1986).
RN [5]
RP SEQUENCE OF 80-597 FROM N.A.
RC MEDLINE=86025405; PubMed=3840370;
RA Chung L.P., Bentley D.R., Reid K.B.M.;
RT "Molecular cloning and characterization of the cDNA coding for C4b-
RL binding protein, a regulatory protein of the classical pathway of the
RL human complement system.";
RL Biochem. J. 230:133-141(1985).
RN [6]
RP SEQUENCE OF 49-88.
RC MEDLINE=85296001; PubMed=4033666;
RA Chung L.P., Gagnon J., Reid K.B.M.;
RT "Amino acid sequence studies of human C4b-binding protein: N-terminal
RL sequence analysis and alignment of the fragments produced by limited
RL proteolysis with chymotrypsin and the peptides produced by cyanogen
RL bromide treatment.";
RL Mol. Immunol. 22:427-435(1985).
RN [7]
RP ELECTRON MICROSCOPY, 3D-STRUCTURE, AND LIGAND BINDING.
RC MEDLINE=83221615; PubMed=6222381;
RA Dahlback B., Smith C.A., Mueller-Eberhard H.J.;
RT "Visualization of human C4b-binding protein and its complexes with
RL vitamin K-dependent protein S and complement protein C4b.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:3461-3465(1983).
CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S

AND WITH SERUM AMYLOID P COMPONENT.
-1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
OF 3 POSSIBLE SORTS: A 570 kDa COMPLEX OF 7 ALPHA CHAINS AND 1
BETA CHAIN, A 530 kDa HOMOPOLYMER OF ALPHA CHAINS OR A 500 kDa
COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF
THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE
BINDING SITE FOR C4B AT THE END.
-1- TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.
-1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
-1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M31452; AAA36507.1; -;
DR EMBL; M62486; AAA36506.1; -;
DR EMBL; M62475; AAA36506.1; JOINED.
DR EMBL; M62476; AAA36506.1; JOINED.
DR EMBL; M62477; AAA36506.1; JOINED.
DR EMBL; M62478; AAA36506.1; JOINED.
DR EMBL; M62479; AAA36506.1; JOINED.
DR EMBL; M62480; AAA36506.1; JOINED.
DR EMBL; M62481; AAA36506.1; JOINED.
DR EMBL; M62482; AAA36506.1; JOINED.
DR EMBL; M62484; AAA36506.1; JOINED.
DR EMBL; M62485; AAA36506.1; JOINED.
DR EMBL; X07853; CAA30701.1; -;
DR EMBL; X04284; CAB51244.1; -;
DR EMBL; X04296; CAA27839.1; -;
DR EMBL; X02865; CAA26617.1; -;
DR PIR; A33568; NBHUC4.
DR HSP; P10998; 1LVUC.
DR MIM; 120830; -;
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 8.
DR SMART: SM00032; CCP; 8.
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal;
KW Polymorphism.
FT SIGNAL 1 48
FT CHAIN 49 597 C4B-BINDING PROTEIN ALPHA CHAIN.
FT DOMAIN 49 109 SUSHI 1.
FT DOMAIN 112 171 SUSHI 2.
FT DOMAIN 174 235 SUSHI 3.
FT DOMAIN 238 295 SUSHI 4.
FT DOMAIN 298 361 SUSHI 5.
FT DOMAIN 364 423 SUSHI 6.
FT DOMAIN 425 481 SUSHI 7.
FT DOMAIN 483 539 SUSHI 8.
FT DISULFID 50 96 BY SIMILARITY.
FT DISULFID 81 108 BY SIMILARITY.
FT DISULFID 113 154 BY SIMILARITY.
FT DISULFID 140 170 BY SIMILARITY.
FT DISULFID 175 217 BY SIMILARITY.
FT DISULFID 203 234 BY SIMILARITY.
FT DISULFID 239 281 BY SIMILARITY.
FT DISULFID 267 294 BY SIMILARITY.
FT DISULFID 299 348 BY SIMILARITY.
FT DISULFID 332 360 BY SIMILARITY.
FT DISULFID 364 387 BY SIMILARITY.
FT DISULFID 365 387 BY SIMILARITY.
FT DISULFID 399 422 BY SIMILARITY.
FT DISULFID 426 468 BY SIMILARITY.
FT DISULFID 454 480 BY SIMILARITY.
FT DISULFID 484 525 BY SIMILARITY.
FT DISULFID 511 538 BY SIMILARITY.
FT DISULFID 546 565 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT DISULFID 558 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT DISULFID 558

```

FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .).
FT VARIANT 92 92 Q -> T.
FT VARIANT 357 357 /FTId=VAR_001977.
FT VARIANT 473 473 /FTId=VAR_001978.
FT VARIANT 473 473 W -> L (IN DBSNP:1801341).
FT SEQUENCE 597 AA; 67033 MW; 67E03F2EA85A16DD CRC64;

Query Match 20.28; Score 303; DB 1; Length 597;
Best Local Similarity 30.78; Pred. No. 6.4e-18;
Matches 80; Conservative 42; Mismatches 101; Indels 38; Gaps 16;

QY 20 SDQYVPECTQAIYKRCRQY-RSLGNVIMVCRK-GEWVALNPLKCRKPCGHPGDTPEGT 77
DB 67 TETREKGTTLKYKLCPLGQYVRSHTQTLTNSDGEWV-YNTF--CIYKRCRHPGE----- 118
QY 78 FTLTGGNV-----FEYGVKAVYTCNEGYYQLLGEINRECDTD----GWTNDIPICEVVKC 128
DB 119 --LRNGQVEIKTDLFSQIEFSCSEGFLLIGSTTSR-CEVQDRGVGWSHPLPQCEIYVKC 175
QY 129 LPVTAPENKGIYSSAMEPDREYHFGQAVFVCSNGYKTEGDEEMHCSDD----GFWSKEK 184
DB 176 KPPPDIRNGRHSGE-----ENFYAYGFVSQYSCDPFSLGLGHASISCTVENETIGVWRPSP 231
QY 185 PKCIVEISCKSPDVNSPISQ-KIYKENERFOYKCNNGYEYSEGRDAV--C-TESGWRP 240
DB 232 PTCEKITCRKPDVSHGEMVSGFGPIYNYKDTIVFCQKGFVL--RGSSVIHCDADSKWNP 289
QY 241 L-PSCEEKSCDN-PYIPNGDY 259
DB 290 SPPACEPNCSINLPDIPHASW 310

RESULT 5
VCP_VACCV STANDARD; PRT; 263 AA.
AC P10998;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement control protein precursor (VCP) (Secretory protein 35)
DE (Protein C3) (28 kDa protein).
GN C3L.
OS Vaccinia virus (strain WR), and
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10254, 10249;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-37.
RC STRAIN=WR;
RX MEDLINE=88318974; PubMed=3412473;
RA Kotwal G.J., Moss B.;
RT "Vaccinia virus encodes a secretory polypeptide structurally related
RL Nature 335:176-178(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WR;
RX MEDLINE=89073756; PubMed=2849238;
RA Kotwal G.J., Moss B.;
RT "Analysis of a large cluster of nonessential genes deleted from a
RL vaccinia virus terminal transposition mutant.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=COPENHAGEN;
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
```

```

RA Paoletti E.;
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [4]
RP COMPLETE GENOME.
RC STRAIN=COPENHAGEN;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RT "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL Virology 179:517-563(1990).
RN [5]
RP FUNCTION.
RX MEDLINE=92115714; PubMed=1731333;
RA Isaacs S.N., Kotwal G.J., Moss B.;
RT "Vaccinia virus complement-control protein prevents
RT antibody-dependent complement-enhanced neutralization of infectivity
RT and contributes to virulence.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:628-632(1992).
RN [6]
RP STRUCTURE BY NMR OF 146-263.
RX MEDLINE=97446168; PubMed=9299352;
RA Wiles A.P., Shaw G., Bright J., Perczel A., Campbell I.D.,
RA Barlow P.N.;
RT "NMR studies of a viral protein that mimics the regulators of
RT complement activation.";
RL J. Mol. Biol. 272:253-265(1997).
CC -I- FUNCTION: SERVES TO PROTECT THE VIRUS AGAINST COMPLEMENT ATTACK BY
CC INHIBITING BOTH CLASSICAL AND ALTERNATIVE PATHWAYS OF COMPLEMENT
CC ACTIVATION. BINDS C3B AND C4B.
CC -I- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF
CC COMPLEMENT ACTIVATION (RCA).
CC -I- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X13166; CAA31564.1; -;
CC EMBL; M22812; AAA69605.1; -;
CC EMBL; M35027; AAA47997.1; -;
CC PIR; A31005; WMVZSP.
CC PDB; 1VVC; 03-DEC-97.
CC PDB; 1VVD; 03-DEC-97.
CC PDB; 1VVE; 03-DEC-97.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00084; sushi; 4.
CC SMART; SM00032; CCP; 4.
CC Signal; Repeat; Sushi; 3D-structure.
CC SIGNAL 1 19
CC CHAIN 20 263 COMPLEMENT CONTROL PROTEIN.
CC DOMAIN 20 82 SUSHI 1.
CC DOMAIN 85 144 SUSHI 2.
CC DOMAIN 147 202 SUSHI 3.
CC DOMAIN 205 262 SUSHI 4.
CC DISULFID 21 70 BY SIMILARITY.
CC DISULFID 54 81 BY SIMILARITY.
CC DISULFID 86 126 BY SIMILARITY.
CC DISULFID 112 143 BY SIMILARITY.
CC DISULFID 148 190 BY SIMILARITY.
CC DISULFID 176 201 BY SIMILARITY.
CC DISULFID 206 248 BY SIMILARITY.
CC DISULFID 234 261 BY SIMILARITY.
CC SEQUENCE 263 AA; 84322CC9A6EF8997 CRC64;

Query Match 20.18; Score 301.5; DB 1; Length 263;
Best Local Similarity 33.38; Pred. No. 3.4e-16;
Matches 78; Conservative 29; Mismatches 106; Indels 21; Gaps 12;
```


FT DISULFID 523 549 BY SIMILARITY.
FT DISULFID 554 595 BY SIMILARITY.
FT DISULFID 581 611 BY SIMILARITY.
FT DISULFID 616 665 BY SIMILARITY.
FT DISULFID 645 682 BY SIMILARITY.
FT DISULFID 688 730 BY SIMILARITY.
FT DISULFID 716 743 BY SIMILARITY.
FT DISULFID 747 790 BY SIMILARITY.
FT DISULFID 776 803 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 866 BY SIMILARITY.
FT DISULFID 871 920 BY SIMILARITY.
FT DISULFID 900 937 BY SIMILARITY.
FT DISULFID 943 986 BY SIMILARITY.
FT DISULFID 973 999 BY SIMILARITY.
FT DISULFID 1004 1045 BY SIMILARITY.
FT DISULFID 1031 1061 BY SIMILARITY.
FT DISULFID 1066 1115 BY SIMILARITY.
FT DISULFID 1095 1132 BY SIMILARITY.
FT DISULFID 1138 1180 BY SIMILARITY.
FT DISULFID 1166 1193 BY SIMILARITY.
FT DISULFID 1197 1240 BY SIMILARITY.
FT DISULFID 1226 1253 BY SIMILARITY.
FT DISULFID 1258 1300 BY SIMILARITY.
FT DISULFID 1286 1316 BY SIMILARITY.
FT DISULFID 1321 1370 BY SIMILARITY.
FT DISULFID 1350 1387 BY SIMILARITY.
FT DISULFID 1396 1439 BY SIMILARITY.
FT DISULFID 1426 1452 BY SIMILARITY.
FT DISULFID 1457 1498 BY SIMILARITY.
FT DISULFID 1484 1514 BY SIMILARITY.
FT DISULFID 1519 1568 BY SIMILARITY.
FT DISULFID 1548 1585 BY SIMILARITY.
FT DISULFID 1591 1633 BY SIMILARITY.
FT DISULFID 1619 1646 BY SIMILARITY.
FT DISULFID 1650 1693 BY SIMILARITY.
FT DISULFID 1679 1706 BY SIMILARITY.
FT DISULFID 1711 1753 BY SIMILARITY.
FT DISULFID 1739 1769 BY SIMILARITY.
FT DISULFID 1774 1823 BY SIMILARITY.
FT DISULFID 1803 1840 BY SIMILARITY.
FT DISULFID 1848 1891 BY SIMILARITY.
FT DISULFID 1877 1904 BY SIMILARITY.
FT DISULFID 1909 1952 BY SIMILARITY.
FT DISULFID 1938 1965 BY SIMILARITY.
FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 509 509 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 860 860 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 959 959 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1028 1028 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1152 1152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1310 1310 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1481 1481 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1504 1504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1540 1540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1605 1605 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1763 1763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1908 1908 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2039 AA; 223589 MW; B2FD29B6AD3C5EB7 CRC64;

Query Match 19.9%; Score 298.5; DB 1; Length 2039;
Best Local Similarity 27.0%; Pred. No. 6.1e-17;
Matches 86; Conservative 49; Mismatches 111; Indels 73; Gaps 18;
QY 3 CNE---LPPRRNTEILTGSWSQDTYPEGTOAIYKRCRPGYRSLGNVINVCRKGWVALNPL 59

Db 43 CNAPENLFPARTNLT---DEFEPIGTLYNRCRPGYSGRPFISICLKNKSVMTGAKD- 97
QY 60 RKQKRCRPGHGDTPFTTTLTGNGVNFYGVKAVYTCNEGYYQLLGEINRYREC-----DTDG 115
Db 98 -RCRRKSCRNPPDPVNGMVHVIKG--IQFGSOKYCTKGYRLIGS-SSATCIISGDTVI 153
QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDRYHFQAVRFVCNSG-----YKLEG 168
Db 154 WDNETPICDRIPCGUPPTI-TNGDFISTNRE---NFHYGSVVYTRCNSGSGGRKKVFLVG 209
QY 169 DEEMHC--SDD--GFWSKEKPKC-VEISCKSPDVINGSPIS-QKIYKENERFOYKCNMG 222
Db 210 EPSIVCTSDNDQGVGWSGPAQCIIIPNKCTPPNVENGILVSDNRSLFSLNEVVEFCQPG 269
QY 223 YEYSERGDVCTE-SGWRP-LPSCDE-----KSCDNP 252
Db 270 FVMKGPVRVKQALNKWPELPSCSRVCPQPPDVLHAERTQDKDNFSPGQEVFVSCPEG 329
QY 253 Y-----IPNGDYSP 261
Db 330 YDLRGAASMRCTPOGDWSP 348
RESULT 7
C4BP_RAT
ID C4BP_RAT STANDARD; PRT; 558 AA.
AC Q63514;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C4b-binding protein alpha chain precursor (C4bp).
GN C4BPA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=97166082; PubMed=9013975;
RA Hillarp A., Wiklund H., Thern A., Dahlback B.;
RT "Molecular cloning of rat C4b binding protein alpha- and beta-chains:
RT structural and functional relationships among human, bovine, rabbit,
RT mouse, and rat proteins.";
RL J. Immunol. 158:1315-1323(1997).
CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S
CC AND WITH SERUM AMYLOID P COMPONENT.
CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.
CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; Z50051; CAA90391.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 8.
DR SMART; SM00032; CCP; 8.
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT SIGNAL 1 13 BY SIMILARITY.
FT CHAIN 14 558 C4B-BINDING PROTEIN ALPHA CHAIN.

```
FT DOMAIN 14 73 SUSHI 1.
FT DOMAIN 76 135 SUSHI 2.
FT DOMAIN 138 200 SUSHI 3.
FT DOMAIN 203 259 SUSHI 4.
FT DOMAIN 262 325 SUSHI 5.
FT DOMAIN 328 387 SUSHI 6.
FT DOMAIN 389 444 SUSHI 7.
FT DOMAIN 446 502 SUSHI 8.
FT DISULFID 15 60 BY SIMILARITY.
FT DISULFID 45 72 BY SIMILARITY.
FT DISULFID 77 118 BY SIMILARITY.
FT DISULFID 104 134 BY SIMILARITY.
FT DISULFID 139 182 BY SIMILARITY.
FT DISULFID 168 199 BY SIMILARITY.
FT DISULFID 204 246 BY SIMILARITY.
FT DISULFID 232 258 BY SIMILARITY.
FT DISULFID 263 312 BY SIMILARITY.
FT DISULFID 296 324 BY SIMILARITY.
FT DISULFID 328 351 BY SIMILARITY.
FT DISULFID 329 373 BY SIMILARITY.
FT DISULFID 363 386 BY SIMILARITY.
FT DISULFID 390 431 BY SIMILARITY.
FT DISULFID 417 443 BY SIMILARITY.
FT DISULFID 447 488 BY SIMILARITY.
FT DISULFID 474 501 BY SIMILARITY.
FT DISULFID 509 509 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT DISULFID 521 521 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT CARBOHYD 31 31 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 558 AA; 62266 MW; 592F0C667ED1E3FF CRC64;

Query Match 18.5%; Score 276.5; DB 1; Length 558;
Best Local Similarity 30.2%; Pred. No. 9.5e-16;
Matches 83; Conservative 43; Mismatches 104; Indels 45; Gaps 17;

QY 7 PPRNTEILGWSNDQT-YPEGQAIYKCRPGY-RSLGNVIMVCRK-GEW---VALNPLR 60
DB 17 PPPDLPYALPASEMNOTDFESHTTLRLVNCRPGYSRASSQSGLCKPLGKQWQINIA----- 71
QY 61 KCKRPGCHGPDTPFGTFTLTGNV-----FEYGVKAVYTCNEGYYQLLG-EINRECDTD 114
DB 72 -CVKSCRNPGD-----LQNGKVEVKTFDLFGSQEFSCEGYILIGSSTSYCEIQGK 123
QY 115 G--WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDRE---YHFGQAVRFVCSNGYKIEG 168
DB 124 GVNSDPLPECVIAKCMGP---PD-----ISNGKHNGREEFFTYRSVTKCPDFTLLG 176
QY 169 DEEMHCS----DGFWSKEKPKVEISCKSPDVINGSPIQ-KIIYKENERFQYKCNMGY 223
DB 177 NASITCTVVKTKVGVNSPSPPTCERIIICPWPVKLHGHTINSGRKHTYKYKDSVRFVCQKG 236
QY 224 EYSEKGDVCTESG-WRPLPSCEKSC-DNPIYPN 256
DB 237 VLRGSGVHCADGWSVPVPCVLSNCTDIPDIPN 271

RESULT 8
C4BP_MOUSE
ID C4BP_MOUSE STANDARD; PRT: 469 AA.
AC P08607;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C4b-binding protein precursor (C4bp).
GN C4BPA OR C4BP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```
RN SEQUENCE FROM N.A.
RP MEDLINE=88024997; PubMed=3663616;
FX Kristensen T., Ogata R.T., Chung L.P., Reid K.B.M., Tack B.F.;
RA *cDNA structure of murine C4b-binding protein, a regulatory component
RT of the serum complement system.*;
RL Biochemistry 26:4668-4674(1987).
CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC (C3BINA). WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
CC -1- SUBUNIT: HOMOHEPTAMER; NOT COVALENTLY LINKED. MOUSE LACKS THE
CC BETA CHAIN OF C4BP.
CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -1- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-44 IS THE INITIATOR.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to a license@isb-sib.ch).
CC -----
CC EMBL; M17122; AAA37312.1; ALT_INIT.
CC PIR; A27117; NBMSC4.
CC HSP; P10998; IVD.
CC MGD; MG1:88229; C4bp.
CC InterPro; IPR000436; Sush1_SCR_CCP.
CC Pfam; PF00084; sush1; 6.
CC SMART; SM00032; CCP; 6.
CC Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT SIGNAL 1 56
FT CHAIN 57 469 C4B-BINDING PROTEIN.
FT DOMAIN 57 116 SUSHI 1.
FT DOMAIN 119 177 SUSHI 2.
FT DOMAIN 180 241 SUSHI 3.
FT DOMAIN 244 300 SUSHI 4.
FT DOMAIN 302 356 SUSHI 5.
FT DOMAIN 358 414 SUSHI 6.
FT DISULFID 58 103 BY SIMILARITY.
FT DISULFID 88 115 BY SIMILARITY.
FT DISULFID 120 160 BY SIMILARITY.
FT DISULFID 146 176 BY SIMILARITY.
FT DISULFID 181 223 BY SIMILARITY.
FT DISULFID 209 240 BY SIMILARITY.
FT DISULFID 245 287 BY SIMILARITY.
FT DISULFID 273 299 BY SIMILARITY.
FT DISULFID 303 343 BY SIMILARITY.
FT DISULFID 329 355 BY SIMILARITY.
FT DISULFID 359 400 BY SIMILARITY.
FT DISULFID 386 413 BY SIMILARITY.
FT CARBOHYD 74 74 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 428 428 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 469 AA; 51551 MW; 41E137CB8D8C6321 CRC64;

Query Match 18.4%; Score 275; DB 1; Length 469;
Best Local Similarity 28.8%; Pred. No. 1e-15;
Matches 78; Conservative 46; Mismatches 119; Indels 28; Gaps 15;

QY 7 PPRNTEILGWSNDQT-YPEGQAIYKCRPGY-RSLGNVIMVCR-KGEW-VALNPLRKC 62
DB 60 PPPAIPNALPASDVNRDTFESHTTLKTECLPGYGRGIRMMVYKPSGEWEISVS----C 115
```


63 QKRPCGHGDTPTFTLTGCVNFEYGVKAVYTCNCGYQLLGEINRECDTDG----WTN 118
116 AKKCRNPGYLDNG--YVNGEITTFQSGQIEFSCQEGFVLVGS-STSSCEVRKGVAVSN 171
119 DIPICEVVKLPVTPANGKIVSAMPEPDREYHFGQAVRFVCSGYKIEGDEMHCS--- 175
172 PFPECVIVKCGPPDISNGK--HSGTDFYPYNHG--ISYTCDFGRFLVGSFPFGCVVN 227
176 -DGFWSKEKPKVCIEISCKSPDVINGSPIS-OKIIYKENERFOYKCNMGVEYSERGDVAVC 233
228 KTVPMVSSSPTECKEIKCSOPNHLGVISGYKATYTHRSVRLACLNGTVLGRHRHVIC 287
234 TEGS-WRLPSCSEKSCD-NPYIPNGDYSPL 262
288 QGNGNWSLPTC-EFDCDLPALVNGYVTSM 317

RESULT 9
MCP_HUMAN STANDARD; PRT; 377 AA.
AC P15529;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Membrane cofactor protein precursor (CD46 antigen) (Trophoblast
leucocyte common antigen) (TLX).
GN Homo sapiens (Human).
OS Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.
RX MEDLINE=88286080; PubMed=3260937;
RA Lublin D.M., Liszewski M.K., Post T.W., Arce M.A., le Beau M.M.,
Rebentisch M.B., Lemons R.S., Seta T., Atkinson J.P.;
RT "Molecular cloning and chromosomal localization of the human membrane
cofactor protein (MCP). Evidence for inclusion in the multigene
family of complement-regulatory proteins.";
RL J. Exp. Med. 168:181-194(1988).
RN [2]
SEQUENCE FROM N.A.
RN TISSUE=Testis;
RX MEDLINE=93119658; PubMed=8418811;
RA Cervoni F., Fenichel P., Akhoundi C., Hsi B.L., Rossi B.;
RT "Characterization of a cDNA clone coding for human testis membrane
cofactor protein (MCP, CD46).";
RL Mol. Reprod. Dev. 34:107-113(1993).
RN [3]
SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=91267562; PubMed=2050389;
RA Purcell D.F., Russell S.M., Deacon N.J., Brown M.A., Hooker D.J.,
McKenzie I.F.;
RT "Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a
regulator of complement activation.";
RL Immunogenetics 33:335-344(1991).
RN [4]
SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=94014356; PubMed=7691939;
RA Cui W., Hourcade D., Post T., Greenlund A.C., Atkinson J.P.,
Kumar V.;
RT "Characterization of the promoter region of the membrane cofactor
protein (CD46) gene of the human complement system and comparison to
a membrane cofactor protein-like genetic element.";
RL J. Immunol. 151:4137-4146(1993).
RN [5]
ALTERNATIVE SPLICING.
RX MEDLINE=92289809; PubMed=1601037;
RA Russell S.M., Sparrow R.L., McKenzie I.F.C., Purcell D.F.J.;
RT "Tissue-specific and allelic expression of the complement regulator
CD46 is controlled by alternative splicing.";
RL Eur. J. Immunol. 22:1513-1518(1992).
CC !- FUNCTION: MCP IS AN IMPORTANT MEMBRANE PROTEIN FOR PROTECTING HOST

CC CELLS FROM DAMAGE BY COMPLEMENT. MCP POSSESSES COFACTOR ACTIVITY
FOR INACTIVATION OF C3B AND C4B BY SERUM FACTOR I. THIS WOULD
ALLOW MCP TO REGULATE THE COMPLEMENT SYSTEM AT THE LEVEL OF THE C3
CONVERTASE. MAY BE INVOLVED IN THE FUSION OF THE SPERMATOZOON WITH
THE OOCYTE DURING FERTILIZATION. MAY ACT BY PREVENTING COMPLEMENT
ACTIVATION AT THE FETO-MATERNAL INTERFACE ON THE
SYNCYTIOFROPHBLAST LAYER OF PLACENTA.
CC !- SUBCELLULAR LOCATION: Type I membrane protein.
CC !- ALTERNATIVE PRODUCTS: AT LEAST 14 FORMS ARE PRODUCED BY
ALTERNATIVE SPLICING. THE FORM SHOWN IS THE TRANSCRIPT A. THE
GAMMA ISOFORM (TRANSCRIPTS A AND B) IS PREFERENTIALLY EXPRESSED IN
EBV-B CELLS AND LEUKEMIC CELLS. THE 66 KDa ALPHA ISOFORM AND THE
56 KDa BETA ISOFORM FOUND IN ALL TISSUES EXCEPT SPERM CORRESPOND
TO THE TRANSCRIPTS C AND D, AND THE TRANSCRIPTS E AND F
RESPECTIVELY. THE EXON 9 IS SPECIFICALLY DELETED IN SOME
PLACENTAE. THE EPSILON ISOFORM CORRESPONDS TO THE TRANSCRIPTS I
AND J. SPERMATOZOON DELTA ISOFORM (TRANSCRIPT N) DELETE BOTH EXONS
12 AND 13. ALL TISSUES DIFFERENTIALLY SPLICE EXON 13.
CC !- TISSUE SPECIFICITY: EXPRESSED BY ALMOST ALL CELLS EXCEPT
ERYTHROCYTES AND SOME BONE MARROW CELLS.
CC !- PTM: EXTENSIVELY O-GLYCOSYLATED IN THE SER/THR-RICH DOMAIN
(PROBABLE).
CC !- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC !- DATABASE: NAME=PROW; NOTE-CD guide CD46 entry;
WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd46.htm".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

CC EMBL: Y00651; CAA68675.1; -
CC EMBL: S51940; AAB24802.1; -
CC EMBL: M58050; AAG2833.1; -
CC EMBL: A18585; CAA01400.1; -
CC EMBL: S65879; AAD13968.1; -
CC PIR: S01896; S01896.
CC HSSP: P10998; 1VVD.
CC MIM: 120920; -
CC InterPro: IPR000436; Sushi_SCR_CCP.
CC Pfam: PF00084; sushi; 4.
CC SMART: SM00032; CCP; 4.
CC Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW Sushi; Alternative splicing.
FT SIGNAL 1 34
FT CHAIN 35 377 MEMBRANE COFACTOR PROTEIN.
FT DOMAIN 35 328 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 329 351 POTENTIAL.
FT DOMAIN 352 377 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 95
FT DOMAIN 98 158 SUSHI 1.
FT DOMAIN 161 224 SUSHI 2.
FT DOMAIN 227 284 SUSHI 3.
FT DOMAIN 287 311 SER/THR-RICH.
FT DISULFID 35 80 BY SIMILARITY.
FT DISULFID 64 94 BY SIMILARITY.
FT DISULFID 99 141 BY SIMILARITY.
FT DISULFID 127 157 BY SIMILARITY.
FT DISULFID 162 210 BY SIMILARITY.
FT DISULFID 191 223 BY SIMILARITY.
FT DISULFID 228 270 BY SIMILARITY.
FT DISULFID 256 283 BY SIMILARITY.
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 286 300 MISSING (IN A SECOND FORM).
FT VARSPLIC 340 352 MISSING (IN ISOFORM M).
FT VARSPLIC 353 361 QSQRRKKK -> DIPKGGRRKKQKQVLMPLTRUNQPLQ
YSQREA (IN ISOFORM M).
FT VARSPLIC 362 377 TYLTDETHREVKFTSL -> KADGGAETATYQTKSTPAEQ


```

FT          /FTid=VAR_011794.
SQ SEQUENCE 610 AA; 66655 MW; 7D43E3C0D1229229 CRG64;

Query Match      18.2%; Score 272; DB 1; Length 610;
Best Local Similarity 27.1%; Pred. No. 2.5e-15;
Matches 70; Conservative 37; Mismatches 107; Indels 44; Gaps

Qy 23  TYPEQTAIVKCRPGYRSLGNVIMVC-RKGEWVALNPLRK---CQ--KRP-----CGHP 70
Db 261  SFPNNTTCTDCEBGFELMAQASLOCTSSGNWNEKPTCAVTCRAVRQPNQSGVRCSH-319
Qy 71  GDTPTGFTTLTGGNVFYGVKAVYTCNMGYQLLGEINRYECDTDG-WTNDIPICEYVKCL 129
Db 320  --SPAGEFT-----FKSSCNFTCEBGFMLQPAQV-ECTTQGWQTQIPVCEAFQCT 368
Qy 130  PVTAPENKIVSSNAMEPDREYHFHQAVRFVNCNYSKYIEGDEEMHCSDDGFWSKEPKCIVE 189
Db 369  ALSNPERG-YMNCPLPSAGSFYRSGSCFSCQGFVLKSKRLQCGPTGEMDNKPTCEA 427
Qy 190  ISCKSPDVIN-----GSPISOKIIVKENERFYQKNCMYEYSERGDAVCTESG-W 238
Db 428  VRC---DAVHQPPKGLVRCASHPIGE-FTYKSSCAF--SCEEGFELHGSTQLECTSQGW 481
Qy 239  -RPLPSCSEKSCDNPYIP 255
Db 482  TEEVPSCQVVKCSSLAVP 499

RESULT 11
ID LEW2_PIG STANDARD; PRT; 484 AA.
AC P98110;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE E-selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE (CD62E).
GN SELE.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OC NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Aortic endothelium;
RX MEDLINE=95071392; PubMed=7526854;
RA Rollins S.A., Evans M.J., Johnson K.K., Elliot E.A., Squinto S.P.,
RA Matis L.A., Rother R.P.;
RT "Molecular and functional analysis of porcine E-selectin reveals a
RT potential role in xenograft rejection.;"
RN Biochem. Biophys. Res. Commun. 204:763-771(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Aortic endothelium;
RX MEDLINE=94271236; PubMed=7516159;
RA Tsang Y.T.M., Haskard D.O., Robinson M.K.;
RT "Cloning and expression kinetics of porcine vascular cell adhesion
RT molecule.;"
RN Biochem. Biophys. Res. Commun. 201:805-805(1994).
CC -1- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC ELAM-1 IS SIALLYL-LEWIS X (ALPHA(1->3)FUCOSYLATED DERIVATIVES OF
CC POLYGLUTAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC GLYCOPIDS).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN ACUTE CELLULAR ALLOGRAFT
CC REJECTION AND PROBABLY ALSO IN XENOGRAFT REJECTION.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: TO OTHER SELECTINS/LECAMs.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 4 SUSL (SCD) DOMAINS.

```



```
FT DOMAIN 173 235 SUSHI 3.
FT DOMAIN 238 295 SUSHI 4.
FT DOMAIN 298 363 SUSHI 5.
FT DOMAIN 366 426 SUSHI 6.
FT DOMAIN 428 484 SUSHI 7.
FT DOMAIN 486 542 SUSHI 8.
FT DISULFID 50 95 BY SIMILARITY.
FT DISULFID 80 107 BY SIMILARITY.
FT DISULFID 112 153 BY SIMILARITY.
FT DISULFID 139 169 BY SIMILARITY.
FT DISULFID 174 217 BY SIMILARITY.
FT DISULFID 203 234 BY SIMILARITY.
FT DISULFID 239 281 BY SIMILARITY.
FT DISULFID 267 294 BY SIMILARITY.
FT DISULFID 299 350 BY SIMILARITY.
FT DISULFID 334 362 BY SIMILARITY.
FT DISULFID 366 390 BY SIMILARITY.
FT DISULFID 367 412 BY SIMILARITY.
FT DISULFID 402 425 BY SIMILARITY.
FT DISULFID 429 471 BY SIMILARITY.
FT DISULFID 457 483 BY SIMILARITY.
FT DISULFID 487 528 BY SIMILARITY.
FT DISULFID 514 541 BY SIMILARITY.
FT DISULFID 549 561 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT DISULFID 561 561 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 525 525 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 610 AA; 68886 MW; D806B270E8A06B58 CRC64;

Query Match 17.7%; Score 265.5; DB 1; Length 610;
Best Local Similarity 27.5%; Pred. No. 8.7e-15;
Matches 76; Conservative 45; Mismatches 110; Indels 45; Gaps 16;

QY 6 LPRRNTTEILTSWSDQTYPEGTQAIKCRPGYR--SLGNVIMVCRKGEWALNPKQ 63
DB 52 IPPYDFAPFINELNETRFETGTLLRYTCRPGYRISRNKFLICDGTDNW---KYREFCV 108
QY 64 KRPCGHPGDPFTFTTGTGNV-----FEYGVRAVTCNEGYYOLLGEIN-YRECDTDG-- 115
DB 109 KRRCENPGE-----LLNGQVIVKTDYSFGSEIEFSCSEGYVLISANSYCOLQDKGVV 161
QY 116 WTNDIFICEVVKLPVAPENGKIVSAMEPDREYHFGQAVREVCNSGYKIEGDEMHCS 175
DB 162 WSDPLPQCIIAKCEPPPTTISNGR--HNGGDED-FYTYGSSVTYSCDRDFSMCLKASISCR 218
QY 176 DD----GEWSKEKPKCVSEISKSPDVINGSPISQKI-----IYKENERFOYKCNMGYEV 225
DB 219 VENKTIQVNSPSPCKKVICVOPVVDG-----KITSFGPIITYQOISIVYACNKGFR 273
QY 226 SERGDV--C-TESGWR-PLPSCSEKSC-DNPVYPN 256
DB 274 --EGDSLHICADSNWPNPPPTCELNGCLGLPHIP 307

RESULT 13
CCPH_HSVSA
ID CCPH_HSVSA STANDARD; PRT: 360 AA.
AC Q01016;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE Complement control protein homolog precursor (CCPH).
GN 4 OR CCPH.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92333688; PubMed=1321287;
```

```
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Blesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058(1992).
RN [2]
RP SIMILARITY TO CCP.
RX MEDLINE=92260674; PubMed=1316492;
RA Albrecht J.-C., Fleckenstein B.;
RT "New member of the multigene family of complement control proteins in
herpesvirus saimiri.";
RL J. Virol. 66:3937-3940(1992).
CC -|- ALTERNATIVE PRODUCTS: A MEMBRANE-BOUND FORM AND A SECRETED FORM
ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC -|- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF
COMPLEMENT ACTIVATION (RCA).
CC -|- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64346; CAA45626.1; -
DR EMBL; X64346; CAA45627.1; -
DR EMBL; X60283; CAA42823.1; -
DR EMBL; X60283; CAA42822.1; -
DR PIR; B42534; WMBE2E.
DR PIR; A42534; WMBE1E.
DR PIR; S24567; S24567.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
KW Signal; Repeat; SUSHI; Transmembrane; Alternative splicing;
KW Glycoprotein.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 360 COMPLEMENT CONTROL PROTEIN HOMOLOG.
FT DOMAIN 83 143 SUSHI 1.
FT DOMAIN 146 206 SUSHI 2.
FT DOMAIN 209 265 SUSHI 3.
FT TRANSMEM 328 350 POTENTIAL.
FT DISULFID 84 125 BY SIMILARITY.
FT DISULFID 111 142 BY SIMILARITY.
FT DISULFID 147 191 BY SIMILARITY.
FT DISULFID 175 205 BY SIMILARITY.
FT DISULFID 210 252 BY SIMILARITY.
FT DISULFID 238 264 BY SIMILARITY.
FT CARBOHYD 36 36 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 289 302 RICNGNCTSMPTQ -> AECACPGSNYPIS (IN
SHORT ISOFORM).
FT VARSPLIC 303 360 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 360 AA; 40006 MW; 6278A6C2ECD49669 CRC64;

Query Match 17.7%; Score 264.5; DB 1; Length 360;
Best Local Similarity 30.6%; Pred. No. 5.8e-15;
Matches 76; Conservative 22; Mismatches 119; Indels 31; Gaps 11;

QY 24 YPEGTOAIYKCRPGYRSLGNVIMVCRKGEWALNPKQKRCQKPGHPTGTFTLTGG 83
DB 44 YPNGTTLHVTCTREGYAKRPVQVTCVNGWTV---PKCKQKKKSTPQDLLNGRYTIVT-G 99
QY 84 NVFEVGVKAVYTCNEGYYQLLGEIN-----YRECDTDGWTNDIPICEVVKLPVAPENGKI 139
DB 139 NVFEVGVKAVYTCNEGYYQLLGEIN-----YRECDTDGWTNDIPICEVVKLPVAPENGKI 139
```

Db 100 NLX-YGSVITTCNSGYSLIGSTTSACLLKRGGRVDWTPRPICDIKKCKPPQIANG-- 156
 QY 140 VSSAMEPDREYHFGQAVRFVCNSGYK--IRGDEMHCSDDGFV-SREKPKCVELCKSPD 196
 Db 157 --THTVWKDYTYLDTPTVYSCNDETCLTLGSPSSKLCSETGSGWPNGETKCEFIKLPQ 214
 QY 197 VINGSPISQKIYKENERFOY-----KCNMGYYSERGDVACTESQWRP-LPSCSEKSCD 250
 Db 215 VAN----AYVEVRKSATSMQYLHINVKYKGFMLYGETPNTCNHGYWSPAIPCEMKISS- 269
 QY 251 NPVIPNGD 258
 Db 270 ----PKGD 273

RESULT 14
 DAF_HUMAN STANDARD; PRT; 381 AA.
 AC P08174; P09679; P78361;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Complement decay-accelerating factor precursor (CD55 antigen).
 GN DAF OR CR OR CD55.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RX MEDLINE=87115845; PubMed=2433596;
 RA Caras I.W., Davitz M.A., Rhee L., Weddell G., Martin D.W. Jr.,
 RA Nussenzweig V.;
 RT "Cloning of decay-accelerating factor suggests novel use of splicing
 RT to generate two proteins.";
 RL Nature 325:545-549(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cervix;
 RA Strausberg R.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 6-381 FROM N.A. (ISOFORM 2).
 RX MEDLINE=87175602; PubMed=2436222;
 RA Medof M.E., Lublin D.M., Hollers V.M., Ayers D.J., Getty R.R.,
 RA Leykam J.F., Atkinson J.P., Tykocinski M.L.;
 RT "Cloning and characterization of cDNAs encoding the complete sequence
 RT of decay-accelerating factor of human complement.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:2007-2011(1987).
 RN [4]
 RP SEQUENCE OF 35-381 FROM N.A. (ISOFORM 2).
 RC TISSUE=Hippocampus;
 RA Kumar V.B., Hyung C., Nakra R., Walters M., Sasser T., Bernardo A.;
 RT "Decay-accelerating factor (DAF; CD 55) in the brain of Alzheimer's
 RT disease patients";
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 1-100 FROM N.A.
 RX MEDLINE=91271256; PubMed=1711208;
 RA Ewlonu U.K., Ravi L., Medof M.E.;
 RT "Characterization of the decay-accelerating factor gene promoter
 RT region.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4675-4679(1991).
 RN [6]
 RP SEQUENCE OF 35-46.
 RC TISSUE=Urine;
 RX MEDLINE=91291869; PubMed=1712233;
 RA Nakano Y., Suigita Y., Ishikawa Y., Choi N.-H., Tobe T., Tomita M.;
 RT "Isolation of two forms of decay-accelerating factor (DAF) from human
 RT urine.";
 RL Biochim. Biophys. Acta 1074:326-330(1991).
 RN [7]
 RP GPI-ANCHOR.

RA MEDLINE=91093238; PubMed=1824699;
 RX Moran P., Raab H., Kohr W.J., Caras I.W.;
 RT "Glycophospholipid membrane anchor attachment. Molecular analysis of
 RT the cleavage/attachment site.";
 RL J. Biol. Chem. 266:1250-1257(1991).
 RN [8]
 RP DISULFIDE BONDS IN SUSHI DOMAINS.
 RX MEDLINE=92305034; PubMed=1377029;
 RA Nakano Y., Sumida K., Kikuta N., Miura N.-H., Tobe T., Tomita M.;
 RT "Complete determination of disulfide bonds localized within the short
 RT consensus repeat units of decay accelerating factor (CD55 antigen).";
 RL Biochim. Biophys. Acta 1116:235-240(1992).
 RN [9]
 RP FUNCTION AS A ECHOVIRUS RECEPTOR.
 RX MEDLINE=95045399; PubMed=7525274;
 RA Ward T., Pipkin P.A., Clarkson N.A., Stone D.M., Minor P.D.,
 RA Almond J.W.;
 RT "Decay-accelerating factor CD55 is identified as the receptor for
 RT echovirus 7 using CELICS, a rapid immuno-focal cloning method.";
 RL EMBO J. 13:5070-5074(1994).
 RN [10]
 RP VARIANT BLOOD GROUP DR(A-).
 RX MEDLINE=94325573; PubMed=7519480;
 RA Lublin D.M., Wallinson G., Poole J., Reid M.E., Thompson E.S.,
 RA Ferdinand B.R., Telen M.J., Anstee D.J., Tanner M.J.A.;
 RT "Molecular basis of reduced or absent expression of
 RT decay-accelerating factor in Cromer blood group phenotypes.";
 RL Blood 84:1276-1282(1994).
 CC -1- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT
 CC CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT
 CC C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.
 CC INTERACTION OF DAF WITH CELL-ASSOCIATED C4B AND C3B POLYPEPTIDES
 CC INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND
 CC FACTOR B TO ENZYMATICALLY ACTIVE C2A AND BB AND THEREBY PREVENTS
 CC THE FORMATION OF C4B2A AND C3BBB, THE AMPLIFICATION CONVERTASES OF
 CC THE COMPLEMENT CASCADE.
 CC -1- FUNCTION: ALSO ACTS AS THE RECEPTOR FOR ECHOVIRUS 7 AND RELATED
 CC VIRUSES (ECHOVIRUSES 13, 21, 29 AND 33).
 CC -1- SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT
 CC HOMODIMER (MINOR FORM).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1/DAF-1 AND 2/DAF-2 (SHOWN
 CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED ON THE PLASMA MEMBRANES OF ALL CELL
 CC TYPES THAT ARE IN INTIMATE CONTACT WITH PLASMA MEMBRANES OF ALL CELL
 CC PROTEINS. IT IS ALSO FOUND ON THE SURFACES OF EPITHELIAL CELLS
 CC LINING EXTRACELLULAR COMPARTMENTS, AND VARIANTS OF THE MOLECULE
 CC ARE PRESENT IN BODY FLUIDS AND IN EXTRACELLULAR MATRIX.
 CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
 CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
 CC ACTIVE SITE ON SCR3 (BY SIMILARITY).
 CC -1- PTM: THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.
 CC -1- POLYMORPHISM: DAF IS RESPONSIBLE FOR THE CROMER BLOOD GROUP
 CC SYSTEM. IT CONSISTS OF AT LEAST SEVEN HIGH-INCIDENCE (CR(A),
 CC TC(A), DR(A), ES(A), WES(B), UMC, AND IFC) AND LOW-INCIDENCE
 CC (TC(B), TC(C), AND WES(A)) ANTIGENS THAT RESIDE ON DAF. IN THE
 CC CROMER PHENOTYPES DR(A-) AND INAB THERE IS REDUCED OR ABSENT
 CC EXPRESSION OF DAF, RESPECTIVELY. IN THE CASE OF THE DR(A-)
 CC PHENOTYPE, A SINGLE NUCLEOTIDE SUBSTITUTION WITHIN EXON 5 ACCOUNTS
 CC FOR TWO CHANGES: A SIMPLE AMINO ACID SUBSTITUTION THAT IS THE
 CC BASIS OF THE ANTIGENIC VARIATION, AND AN ALTERNATIVE SPLICING
 CC EVENT THAT UNDERLIES THE DECREASED EXPRESSION OF DAF IN THIS
 CC PHENOTYPE.
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
 CC (RCA) FAMILY.
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD55 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cg/cd55.htm".
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way

DR	PFam; PF00008; EGF; 1.	
DR	PFam; PF00059; lectin_c; 1.	
DR	PFam; PF00084; sushi; 8.	
DR	PRINTS; PR00343; SELECTIN.	
DR	SMART; SM00032; CCP; 8.	
DR	SMART; SM00034; CLECT; 1.	
DR	SMART; SM00181; EGF; 1.	
DR	PROSITE; PS00022; EGF_1; 1.	
DR	PROSITE; PS01186; EGF_2; 1.	
DR	PROSITE; PS00615; C-TYPE_LECTIN_1; 1.	
DR	PROSITE; PS50041; C-TYPE_LECTIN_2; 1.	
KW	Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;	
KW	Selectin; Signal; Sushi; Repeat; Lipoprotein; Palmitate.	
FT	SIGNAL	1
FT	CHAIN	41
FT	CHAIN	42 768
FT	DOMAIN	42 709
FT	TRANSMEM	710 733
FT	DOMAIN	734 768
FT	DOMAIN	58 158
FT	DOMAIN	159 195
FT	DOMAIN	199 258
FT	DOMAIN	261 320
FT	DOMAIN	323 382
FT	DOMAIN	385 444
FT	DOMAIN	447 506
FT	DOMAIN	509 568
FT	DOMAIN	579 638
FT	DOMAIN	641 700
FT	DISULFID	60 158
FT	DISULFID	131 150
FT	DISULFID	163 174
FT	DISULFID	168 183
FT	DISULFID	185 194
FT	DISULFID	200 244
FT	DISULFID	230 257
FT	DISULFID	262 306
FT	DISULFID	292 319
FT	DISULFID	324 368
FT	DISULFID	354 381
FT	DISULFID	386 430
FT	DISULFID	416 443
FT	DISULFID	448 492
FT	DISULFID	478 505
FT	DISULFID	510 554
FT	DISULFID	540 567
FT	DISULFID	580 624
FT	DISULFID	610 637
FT	DISULFID	642 686
FT	DISULFID	672 699
FT	CARBOHYD	398 398
FT	CARBOHYD	603 603
FT	CARBOHYD	654 654
FT	CARBOHYD	661 661
FT	CARBOHYD	679 679
FT	LIPID	745 745
FT	SITE	756 759
FT	CONFLICT	724 724
SQ	SEQUENCE	768 AA; 83098 MW; E5173074D2F66868 CRC64;

Db	229	SCAEGYELDGGEL---OC	LAGIWTNNPKCDAVQC	LEAPPHGTM--ACMHP	TA	283
Qy	152	FGQAVRVCNKGIEDE	EMHCSDDGFWSEK	PKCVSEIKSPDV-INGS----	PISOK	206
Db	284	YDSCKEFCQCPGYP	ARGSNLT	LHCTGSCWSEPL	PTCEATACEPPE	343
Qy	207	IYKENERFOYKCM	NGYEYSE	RGDAVCTESG-W-RPLP	SCEEKSCNDNP	256
Db	344	FGY--NSSCFTFL	CAEGVLKGNDA	IOCADSGW	TAPFCEALOC	393

Search completed: August 30, 2002, 06:22:41
Job time: 335 sec

GenCore version 4.5
 Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2002, 06:16:26 ; Search time 77.9 Seconds
 (without alignments)
 588.494 Million cell updates/sec

Title: US-09-316-163-10
 Perfect score: 1497
 Sequence: 1 EDCNELPRRNTTEILGWSWS.....EKSCDNPYPNGDYSPLRIK 265

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : SPTREMBL_19:*
 1: sp_archaea:*
 2: sp_bacteria:*
 3: sp_fungi:*
 4: sp_human:*
 5: sp_invertebrate:*
 6: sp_mammal:*
 7: sp_mhc:*
 8: sp_organelle:*
 9: sp_phage:*
 10: sp_plant:*
 11: sp_rodent:*
 12: sp_virus:*
 13: sp_vertebrate:*
 14: sp_unclassified:*
 15: sp_rvirus:*
 16: sp_bacteriaph:*
 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1497	100.0	1172	4 Q9NU87	Q9nu87 homo sapien
2	1090	72.8	1236	11 Q91YB6	Q9lyb6 rattus norv
3	699	46.7	669	6 Q28085	Q28085 bos taurus
4	440	29.4	1053	13 Q91275	Q91275 paralabrax
5	332.5	22.2	360	12 Q9YTO8	Q9ytq8 ateline her
6	322.5	21.5	550	12 P88903	P88903 kaposi's sa
7	322.5	21.5	550	12 Q40912	Q40912 kaposi's sa
8	320	21.4	645	12 Q9WRU2	Q9wru2 macaca mula
9	310	20.7	259	12 P87616	P87616 cowpox viru
10	308.5	20.6	395	12 Q9J2M6	Q9j2m6 macaca mula
11	308.5	20.6	522	6 Q28769	Q28769 papio cynoc
12	301	20.1	679	11 Q99254	Q99254 mus musculu
13	300.5	20.1	3567	11 Q9ES77	Q9es77 mus musculu
14	298.5	19.9	559	4 Q9UQV2	Q9uqv2 homo sapien
15	298.5	19.9	2039	4 Q16745	Q16745 homo sapien
16	298.5	19.9	2489	4 Q16744	Q16744 homo sapien

ALIGNMENTS

RESULT 1

Q9NU87 ID Q9NU87 PRELIMINARY; PRT; 1172 AA.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. OX NCBI_TaxID=9606; RN [1] SEQUENCE FROM N.A. RP Bird C.; RA Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases. RL EMBL; AL049744; CAB70597.1; -. DR HSSP; P086603; 1HFH. DR InterPro; IPR000436; SushL_SCR_CCP. DR Pfam; PF00084; sush1; 19. DR SMART; SM00032; CCP; 19. SQ SEQUENCE 1172 AA; 132087 MW; 8F5B954C4B4FA454 CRC64;

Query Match 100.0%; Score 1497; DB 4; Length 1172;
 Best Local Similarity 100.0%; Pred. No. 6,9e-135;
 Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDCNELPRRNTTEILGWSWS...DTYPEGTQAIYKCRPGYSLGNVIMVCRGEWALNPLR 60
 |||
 Db 19 EDCNELPRRNTTEILGWSWS...DTYPEGTQAIYKCRPGYSLGNVIMVCRGEWALNPLR 78
 |||
 QY 61 KCQKPCGHPGDPGTFTLTGNVFEYGVKAVYTCNEGYPQLLGEINRECDTGTNDI 120
 |||
 Db 79 KCQKPCGHPGDPGTFTLTGNVFEYGVKAVYTCNEGYPQLLGEINRECDTGTNDI 138
 |||
 QY 121 PTCEVVKLPVTAPENKGTIVSSAMEPDRYHFQGVAFVFCVNSGYKTGDEMKHSCDDGFW 180
 |||

Db 139 PICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVNSGYKIEGDEEMHCSDDGF 198
Qy 181 SKKPKCVETSCSPDVINGSPISQIIYKENERFOYKCNMGVEYSEYSGDAVCTESGWRP 240
Db 199 SKKPKCVETSCSPDVINGSPISQIIYKENERFOYKCNMGVEYSEYSGDAVCTESGWRP 258
Qy 241 LPSCEKSCDNPYPNGDYSPLRIK 265
Db 259 LPSCEKSCDNPYPNGDYSPLRIK 283
RESULT 2
ID Q91YB6 PRELIMINARY; PRT: 1236 AA.
AC Q91YB6;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DE COMPLEMENT INHIBITORY FACTOR H.
GN FH.
OS Rattus norvegicus (Rat).
OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=LIVER;
RA Demberg T., Goetze O., Schlaf G.;
RT "Rat complement factor H: molecular cloning, sequencing and expression
in tissues and isolated cells.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ320522; CAC67513.1;
SQ SEQUENCE 1236 AA; 140343 MW; 1AC89PFA28232EBF CRC64;

Query Match 72.8%; Score 1090; DB 11; Length 1236;
Best Local Similarity 71.3%; Pred. No. 8.4e-96;
Matches 189; Conservative 25; Mismatches 51; Indels 0; Gaps 0;

Qy 1 EDCNELPPRNTTEILTGSWSQDTYPECTQAIYKRCPCYRSILGNVIMVCRKGEMVALNPLR 60
Db 19 EDCKGPPPRENSELSSWSSEQLYSECTQATYKRCPCYRTLTGIVKCKNGEWPSPSR 78
Qy 61 KCKRCPCHPGDTPFGFTTGTGNVFYKAVYTCNEGQYLLGEINRECDTDGWTNDI 120
Db 79 ICKRCPCHPGDTPFGFRLAVGSEFEFGAKVYTCDEGYQLLGEIDYRECDADGWTNDI 138
Qy 121 PICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVNSGYKIEGDEEMHCSDDGF 180
Db 139 PICEVVKCLPVTLENGRIVSGAEPDQYIFGQVRFECNSGFKIEGKEMHCSENGLW 198
Qy 181 SKKPKCVETSCSPDVINGSPISQIIYKENERFOYKCNMGVEYSEYSGDAVCTESGWRP 240
Db 199 SNEKPCQVEISCLPPRVENGDIYLPVYKENERFOYKCKQGFYKRGDAVCTGSGWNP 258
Qy 241 LPSCEKSCDNPYPNGDYSPLRIK 265
Db 259 QPSCCEMTCLTPYIPNGIYTPHRIK 283

RESULT 3
ID Q28085 PRELIMINARY; PRT: 669 AA.
AC Q28085;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE CCP MODULES 3-12, WITH PARTS OF CCP 2 AND 13 (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;

RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=96202005; PubMed=8615824;
RA Soames C.J., Day A.J., Sim R.B.;
RT "Prediction from sequence comparisons of residues of factor H involved
in the interaction with complement component C3b.";
RL Biochem. J. 315:523-531(1996).
DR EMBL; X98697; CAA67257.1;
DR HSP; P10998; LVVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 11.
DR SMART; SM00032; CCP; 11.
FT NON_TER 1
FT TER 669
FT NON_TER 669
SQ SEQUENCE 669 AA; 75683 MW; D0D9DB30EE747AC2 CRC64;

Query Match 46.7%; Score 699; DB 6; Length 669;
Best Local Similarity 63.4%; Pred. No. 1.3e-58;
Matches 118; Conservative 26; Mismatches 42; Indels 0; Gaps 0;

Qy 76 GTFTLTGNNVFYKAVYTCNEGQYLLGEINRECDTDGWTNDIPICEVVKCLPVTAP 135
Db 3 GSPLAEGNQFEYGAQVYTCDEGYQVGMENFRECDTNGWTNDIPICEVVKCLPVTPE 62
Qy 136 NGKIVSSAMEPDREYHFGQAVRFVNSGYKIEGDEEMHCSDDGFWSKEKPKCVETSCSP 195
Db 63 NGKIFSDALEPDQBYTYGQVVFECNSGYMLDGPQIHCSAGGVSAAETPKCVETFCPP 122
Qy 196 DVINGSPIQIIYKENERFOYKCNMGVEYSEYSGDAVCTESGWRPLPSCSEKSCDNPV 255
Db 123 VILNGQAVLPKATYKQNERVQYRCAGFEYQGRGDTYCKSGWTPTCTCITCDPPRP 182
Qy 256 NGDYSP 261
Db 183 NGVYRP 188

RESULT 4
ID Q91275 PRELIMINARY; PRT: 1053 AA.
AC Q91275;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE COMPLEMENT REGULATORY PLASMA PROTEIN.
OS Parabrax nebulifer (barred sand bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Serranidae; Paralabrax.
OX NCBI_TaxID=30873;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=94318039; PubMed=8042982;
RA Dahmen A., Kaldon T., Zipfel P.F., Gigli I.;
RT "Cloning and characterization of a cDNA representing a putative
complement-regulatory plasma protein from barred sand bass (Parablax
nebulifer).";
RL Biochem. J. 301:391-397(1994).
DR EMBL; L21703; AAA92556.1;
DR HSP; P08603; LHFH.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 16.
DR SMART; SM00032; CCP; 16.
SQ SEQUENCE 1053 AA; 117597 MW; F27E32C3AD76D5D3 CRC64;

Query Match 29.4%; Score 440; DB 13; Length 1053;
Best Local Similarity 36.2%; Pred. No. 1.6e-33;
Matches 89; Conservative 35; Mismatches 108; Indels 14; Gaps 6;

```
QY 21 DQTPYEGTOAIYKCRGYSRLGNVIMVCRKGEWALNPLRKKOKRPGCHPGDTPFGTFTL 80
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
47 EASYPGGROVRVCGNVGYS--GFFKLVCVEGKMETRG--AKQCPSCGHPGDAQADFHL 102
QY 81 TGGNVEFYGVKAVYTCNEGVLGELNRECDTGDGNTDIPICEVVKCLPVTAPENGKIV 140
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
103 AEGNDFVSGVYTCOKGQVMYSRINRYRCVAGWGVDVVPVCSQOCLIHVDNNVQVI 162
QY 141 SSAMEPDREYHFGOAVRVCNSGYKI-EGDEEMHCSDDGFWMSKEKPKCIVEISKSPDVIN 199
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
163 GG---PEEATGCVNRFSCSRSEILDGSPELYCDERGDWSGVPKPKCAITCAIPPIEN 218
QY 200 GSPISKIYIKENERFOYKCNMGYSEYSERGDAVCTESG----WRPLPSCEEKSCDNPYP 255
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
219 GNVPGAIRYKENDVLHYECDRAFKHIDR-PSTCIKOGIKAWSPTPLCESIKRLTIMD 277
QY 256 NGDYSP 261
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
278 GTRYEP 283

RESULT 5
QYTO8 PRELIMINARY; PRT; 360 AA.
AC Q9YTO8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE COMPLEMENT CONTROL PROTEIN HOMOLOG CCPH.
OS AteLine herpesvirus 3.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=85618;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=73;
RX MEDLINE=200911363; PubMed=10623770;
RA Albrecht J.C.;
RT "Primary structure of the Herpesvirus Ateles genome.";
RL J. Virol. 74:1033-1037(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=73;
RA Albrecht J.-C., Fleckenstein B.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083424; AAC95530.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
SQ SEQUENCE 360 AA; 40208 MW; 118CF83C034352A0 CRC64;

Query Match 22.2%; Score 332.5; DB 12; Length 360;
Best Local Similarity 34.9%; Pred. No. 9,1e-24;
Matches 96; Conservative 32; Mismatches 106; Indels 41; Gaps 18;

QY 8 PRNTEI-----LTGSWSDQIYPEGTOAIYKCRPGYRSLGNVI--MVCRKGEWALNPL 59
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
24 PKRNVSLRYSLRYNITN--SSGSYPNGTTLQVTCRKGY--IGROIQTVTCVNGNWTVPN-- 77
QY 60 RKCQKRPCGHPGDTFGTFTLGGNVVEYGVKAVYTCNEGYSQYLLGEINRYEC--DTDG-- 115
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
78 -EQARRKSTPADLLNGWTVT-GLNLY-VGSVITYTCNTGYQLLGSPST-SSCLLGPDRGV 133
QY 116 -WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDREYH-FGOAVRVCNSGYK--TEGDEE 171
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
134 NWTPRPICEITKCRPPPTPIANGTHNI-----KEYYYILDVAVTYSQNDETKLTGTGSS 188
QY 172 MHCSDGFW-SKEKPKCIVEISKSPDVINGSPIOSKIYIKENERFOY---KCNMGYEXSE 227
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
189 KQCSETGRWPFDEETKCFKVKIPQVANGHVEVRKT--SNNVQYQYVNIKCDKGFRLQG 246
```

```
QY 228 RGDVACTESGWRP-LPSCEEKS---CDNPYPINGD 258
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
247 ETPNMCKNGWFEALPTCEKPAAPPGRGMDPHDSGE 281

RESULT 6
P88903 PRELIMINARY; PRT; 550 AA.
ID P88903;
AC P88903;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
ORF 4.
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=37296;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97094384; PubMed=8939871;
RA Moore P.S., Boshoff C., Weiss R.A., Chang Y.;
RT "Molecular mimicry of human cytokine and cytokine response pathway genes by KSHV.";
RL Science 274:1739-1744(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97121480; PubMed=8962146;
RA Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D., Party J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RT "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8).";
RL Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D., Party J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U75698; AAC57082.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
SQ SEQUENCE 550 AA; 60688 MW; D4B8B2B4BACD1CB5 CRC64;

Query Match 21.5%; Score 322.5; DB 12; Length 550;
Best Local Similarity 31.2%; Pred. No. 1.4e-22;
Matches 72; Conservative 31; Mismatches 109; Indels 19; Gaps 9;

QY 33 KCRPGYRSLG-NVIMVC-RKGEWALNPLRKKOKRPGCHPGDTPFGTFTLTGG-NVFEY 89
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
52 RCRSGYTVARNITATCLOGGTW--SEPTATCNKSCPNPGEIQNGKVIHGGQDALKYG 109
QY 90 VRAYVTCNEGYSQYLLGEINRYEC-----DYDGTNDIPICEVVKCLPVTAPENGKIVSSAM 144
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
110 ANISVYCNVEYFLVGREYVRYCMIGASGQMAWSSSPFCEKEKC-----HRPKIKNGDF 163
QY 145 EPDREYH-FGOAVRVCNSGYKIEGDEEMHCSDDGFWMSKEKPKCIVEISKSPDVINGSPI 203
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
164 KPDKDYIEYNDVAVHFECEYGLVGHPSITACAVNNTWTSNMPTCELAGCKFFSVTHGYPI 223
QY 204 SQ-KIYIKENERFOYKCNMGYSEYSERGDAVCTESGW-RPLPSCEEKSCDNP 252
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
224 QGFSLYTKHKQSVTFACNDGFLVLRGSPITTCNTVNTWDPLPKCVLEDIDDP 274

RESULT 7
O40912 PRELIMINARY; PRT; 550 AA.
ID O40912;
AC O40912;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
```



```
Query Match      20.7%; Score 310; DB 12; Length 259;
Best Local Similarity 33.5%; Pred. No. 8.7e-22;
Matches 84; Conservative 29; Mismatches 116; Indels 22; Gaps 13;

QY 3 CNELPRRTEILTSGWSQDTYEGTQAIYKCRPGYR--SLGNVIMVCRKGEWALNPLR 60
Db 20 CCPIFSRPTIMFKGT-VDSHNIGDTIYELCLPGYKOKMGPIYAKCTGTGWTLEN--- 75

QY 61 KCQKRCPCGHPGDTPTGFTLTGCVNFYGVKAVYTCNKGQYLLGEI-NYRECDTDG---W 116
Db 76 QCIKRRCPSPRDINDQDLGG---VDFGSSITYSCNSGYHLIGESKSYCELSTGSMW 132

QY 117 TNDIPICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVNCNSGYKIEGDEEMHCS 176
Db 133 NPEAPICESVKQSPPSISNGR--HNGYED--FYTDGSVVITYSCNSGYSLIGNSGYLCS- 187

QY 177 DGFWSKEPKCKVEISCKSPDVINGS-PISQIIYKENERFQYKCNMGYYSERGDVACTE 235
Db 188 GGEWS-PPPTCQIVKCPHTTISNGYLSGFKRSYNDNVDFKCKGYKLSGSSSTCSP 246

QY 236 SG-WRP-LPSC 244
Db 247 GNTWQPELPKC 257

RESULT 10
ID Q9J2M6 PRELIMINARY; PRT; 395 AA.
AC Q9J2M6;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE COMPLEMENT BINDING PROTEIN.
OS Macaca mulatta rhadinovirus 26-95.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=119193;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MACACA MULATTA RHADINOVIRUS ISOLATE 26-95;
RA Alexander L., Denekamp L., Knapp A., Auerbach M.R., Damania B.,
RA Desrosiers R.C.;
RT "The primary sequence of rhesus monkey rhadinovirus isolate 26-95;
RT sequence similarities to Kaposi's sarcoma-associated herpesvirus and
RT rhesus monkey rhadinovirus isolate 17577."
RL J. Virol. 74:3388-3398(2000).
DR EMBL; AF210726; AAF59982.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR001230; Prenyltn.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
SQ SEQUENCE 395 AA; 43922 MW; BAC9C6F2E226AE06 CRC64;

Query Match      20.6%; Score 308.5; DB 12; Length 395;
Best Local Similarity 30.4%; Pred. No. 2e-21;
Matches 76; Conservative 40; Mismatches 113; Indels 21; Gaps 12;

QY 7 PPRRTEILTSGWSQDTYEGTQAIYKCRPG-YRSLGNVIMVCRKGEWALNPLRKQK 64
Db 30 PPFDFRMVKTAN-QNENAVGTGRVRELICRPGFYKIQAVVYECLSNGTWT--TPNAECRR 86

QY 65 RPCGHPGDTPTGFTLT-GGNVFYGVKAVYTCNKGQYLLGEINREC---DTD--GWT 117
Db 87 KRCSNPEDILNGEVIITSDNAFKEGSNITYKCNTRYLLGLG-ATVRTCLLYKYSNLVDWQ 145

QY 118 NDIPICEVVKCLPVTAPENGKIVSSAMEPDRE-YHFGQAVRFVNCNSGYKIEGDEEMHCS 176
Db 146 PAAPTCEIEKCKKQPDIEGKYY-----PVQEFYNYLETITFTCNKDFSLIGNTTTTCWT 200

Query Match      20.6%; Score 308.5; DB 6; Length 522;
Best Local Similarity 29.8%; Pred. No. 2.9e-21;
Matches 84; Conservative 50; Mismatches 111; Indels 37; Gaps 17;

QY 3 CN---ELPPRRNTEILTSGWSQDTYEGTQAIYKCRPGYRSLGNVIMVCRKGEWALNPL 59
Db 48 CNAPEQLPEARPTNLTDS-----EFPVGYLYKECLPGYHGKPFSTICLKNSVMTSAKD- 102

QY 60 RKQKRCPCGHPGDTPTGFTLTGCVNFYGVKAVYTCNKGQYLLGEINRECDDG---- 115
Db 103 -KCTRKSCRNPQDPVNGMYHVI--KDIQFGSQINYSCKNGYRLIGS-SSATCIISGNTVI 158

QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDREYH-FGQAVRFVNCNSG-----YKIE 167
Db 159 WDNETPICIEIIPCGLPPTI-ANGDFISTS----REYFPGYVVITYRNLGSRGKLFELV 213

QY 168 GDEEMHCS--DD--GFWSKEPKCK-VEISCKSPDVINGS-PIS-QKIYKENERFQYKCNM 221
Db 214 GEPSTYCTSKDDQVGIWSPAPQCIIIPNKMPPNVGVLSVYNRSLFSLNVEYFRCOP 273

QY 222 GYEYSERGDVACTE-SGWRP-LPSCSEKSCDNPYPNGDYSP 261
Db 274 GFVMKGPVRVQCALNKWEPELPSCSRVCPPEILLHGEHTP 315

RESULT 12
ID Q99254 PRELIMINARY; PRT; 679 AA.
AC Q99254;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
```

```

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE COMPLEMENT RECEPTOR TYPE 2 (CR2) (FRAGMENT).
GN CR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE=90229754; PubMed=2139460;
RA Kurtz C.B., O'Toole E., Christensen S.M., Weis J.H.;
RT "The murine complement receptor gene family. IV. Alternative splicing
RT of Cr2 gene transcripts predicts two distinct gene products that share
RT homologous domains with both human CR2 and CR1."
RL J. Immunol. 144:3581-3591(1990).
RN [2]
RP SEQUENCE OF 21-367 FROM N.A.
RX MEDLINE=95105691; PubMed=7528766;
RA Kim Y.U., Kinoshita T., Molina H., Hourcade D., Seya T., Wagner L.M.,
RA Holers V.M.;
RT "Mouse complement regulatory protein Crry/p65 uses the specific
RT mechanisms of both human decay-accelerating factor and membrane
RT cofactor protein."
RL J. Exp. Med. 181:151-159(1995).
DR EMBL; U17128; AAA78271.1; -.
DR EMBL; U17123; AAA78271.1; JOINED.
DR EMBL; U17124; AAA78271.1; JOINED.
DR EMBL; U17125; AAA78271.1; JOINED.
DR EMBL; U17126; AAA78271.1; JOINED.
DR EMBL; U17127; AAA78271.1; JOINED.
DR EMBL; M36470; AAA37449.1; -.
DR HSSP; P10998; 1VVD.
DR MGD; MGI:88489; Cr2.
DR InterPro; IPR002396; Selectin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 10.
DR PRINTS; PR00343; SELECTIN.
DR SMART; SM00032; CCP; 10.
FT NON_TER 679
SQ SEQUENCE 679 AA; 74916 MW; 52FC00FDCED20CDC CRC64;

Query Match 20.1%; Score 301; DB 11; Length 679;
Best Local Similarity 26.0%; Pred. No. 2.1e-20;
Matches 82; Conservative 51; Mismatches 94; Indels 88; Gaps 16;

QY 20 SDQT-YPEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPLRKQRCRPGCHGDPFGTF 78
DB 37 SOKSEFAITGTWEYKCRPGYFRKSFITCLTETSKW--SDAQCFKRRKPCMNQPEPLHGSV 94

QY 79 TLTGGNVFEYGVKAVYTCNEGVLGELINVRQC---DTDGNTDIPICEVVKCLPVTAP 134
DB 95 HINTG--IEFGSTITYSCNQYRLIGD--SSATCIVSDNTVMNDNMPLESCIPESPPIAI 151

QY 135 ENKQIVSSAMEPDREYHFGQAVRVCNSG-----YKIEGDEMHCSDD-----GFWSKSK 184
DB 152 SNGDFYSSSRD---SPFYGMVYTYCHTGNREKFLDLVGKSIYCTSKDNQVGIWNSPP 208

QY 185 PKCV-EISCKSPDVG-----SPIS-----204
DB 209 PQCIPRVKCPMEIEENGLVSEFKHSFFLNDVIFCKSGFTMKSGRIAMCPNSKWSPP 268

QY 205 -----QKIY-----KENERF-----QYKCNMGYEYSERGDVCTESG-W-RPL 241
DB 269 LPTCEMGLCPPQNILHDYNNKDERFSVGQKSYTCNPGCYTLIGTNLVCTSLGTWSNTV 328

QY 242 PSCEKSCNDPIPN 256
DB 329 PTCEVKSCD--AIPN 341

RESULT 13

```

```

Q9ES77 PRELIMINARY; PRT; 3567 AA.
ID Q9ES77;
AC Q9ES77;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE POLYDOM PROTEIN PRECURSOR.
GN POLYDOM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C3H/HENSIC;
RX MEDLINE=20517255; PubMed=11062057;
RA Gilges D., Vinit M.-A., Callebaut I., Coulombel L., Cacheux V.,
RA Romeo P.-H., Vigon I.;
RT "Polydom : a secreted protein with pentraxin, complement control
RT protein, epidermal growth factor and von willebrand factor A
RT domains."
RL Biochem. J. 352:49-59(2000).
DR EMBL; AF206329; AAG32160.1; -.
DR HSSP; P00740; 1EDM.
DR MGD; MGI:1928849; Polydom.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR003410; HYR.
DR InterPro; IPR001759; Pentaxin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00008; EGF; 10.
DR Pfam; PF02494; HYR; 2.
DR Pfam; PF00084; sushi; 33.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00895; PENTAXIN.
DR PRINTS; PR00453; VWFADOMAIN.
DR ProDom; PD002153; Pentaxin; 1.
DR SMART; SM00032; CCP; 34.
DR SMART; SM00181; EGF; 15.
DR SMART; SM00179; EGF_Ca; 9.
DR SMART; SM00001; EGF_like; 3.
DR SMART; SM00159; PTX; 1.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
KW Signal.
FT SIGNAL 1 17 POTENTIAL.
SQ SEQUENCE 3567 AA; 387391 MW; 8FBAB276E12293E5 CRC64;

Query Match 20.1%; Score 300.5; DB 11; Length 3567;
Best Local Similarity 29.1%; Pred. No. 1.8e-19;
Matches 73; Conservative 34; Mismatches 119; Indels 25; Gaps 9;

QY 17 GSWSDQTYPEGTQAIYKCRPGYRSLGNVIMVC-RKGEWVALNPLRKQRCRPGCHGDPFP 75
DB 1797 GHSSGIYTVGTAVTETSCDEGHVGLVGVSTITCLTETGWDRLRD--SCEATSCGP---PV 1851

QY 76 GFTLTGGNVFEYGVKAVYTCNEGVLGELINVRQC---DTDGNTDIPICEVVKCLPVTAP 135
DB 1852 PNGGVGDGSAFTYGVKAVYRCKDGYTLSDGESACIASGWSHSSPVCLVKCSQPEDIN 1911

QY 136 NGKIVSSAMEPDREYHFGQAVRVCNSGKYKIEGDEMHCSDDGFWKSKPKVCVEISCKSP 195
DB 1912 NGKYILSLGT-----YLSIASYSCENGYSLQGPSLLETCTAGSWDRAPPSCOLVSCGEP 1965

QY 196 DVINGSPISQKIYKENERF---QYKCNMGYEYSERGDVCTESG-WRPL-PSCEKSC 249
DB 196 DVINGSPISQKIYKENERF---QYKCNMGYEYSERGDVCTESG-WRPL-PSCEKSC 249

```

Db 1966 PIVKDA-----VITGSNFTFGNTVATYCKEGYTLAGPDTIVCQANGKWNSSNHQCLAVSC 2020

QY 250 DNPYIPNDYS 260

Db 2021 DEP--PNVDHA 2029

RESULT 14

Q9UQV2

ID Q9UQV2 PRELIMINARY; PRT; 559 AA.

AC Q9UQV2;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE CR1 C3b/C4b RECEPTOR SECRETED FORM PRECURSOR (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=89010527; PubMed=2971757;

RA Hourcade D., Miesner D.R., Atkinson J.P., Holers V.M.;

RT "Identification of an alternative polyadenylation site in the human

RT C3b/C4b receptor (complement receptor type 1) transcriptional unit and

RT prediction of a secreted form of complement receptor type 1.";

RL J. Exp. Med. 168:1255-1270(1988).

DR EMBL; X14362; CAA32541.1;

DR HSSP; P10998; 1VVD.

DR InterPro; IPR002396; Selectin.

DR InterPro; IPR000436; Sushi_SCR_CCP.

DR InterPro; IPR000834; zn-carbopept.

DR Pfam; PF00084; sushi; 8.

DR PRINTS; PR00343; SELECTIN.

DR SMART; SM00032; CCP; 8.

DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.

KW Signal; Receptor.

FT NON_TER 1

FT SIGNAL <1 16 POTENTIAL.

FT CHAIN 17 559 POTENTIAL.

SQ SEQUENCE 559 AA; 61424 MW; DBFFE965CA179D75 CRC64;

Query Match 19.9%; Score 298.5; DB 4; Length 559;

Best Local Similarity 27.0%; Pred. No. 2.8e-20;

Matches 86; Conservative 49; Mismatches 111; Indels 73; Gaps 18;

QY 3 CNE---LPPRRNTEILTGSWSQTYPEGTQAIYKCRPGYRSLGNVTMVCRCGEWALNPL 59

Db 18 CNAPEWLPARTNLT-----DFEFPIGIYLYNECRPGYSGRPFSLICLKNSVWTGAKD- 72

QY 60 RKCQRKPCGHPGDTFGFTLTGGNVFEYGVKAVYTCNEGYSQLLGEINRYREC----DTDG 115

Db 73 -RCRRKSCRNPDPVNGMVHVIKG--IQFGSIIKYSCTKGRLIGS-SSATCLISGDTVI 128

QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEDREYHFCQAVFVCSNG-----YKIEG 168

Db 129 WDNETPICDIRPCGLPPTT-TGDFISTNRE---NFHYGSVYTYRCNPGSGGRKVFELVG 184

QY 169 DEEMHC--SDD--GEWSKPKPC-VEISCKSPDVLINGSPIIS-OKIYKENERFOYKCNKG 222

Db 185 EFSIVCTSDNDQGVITSGPAPCIIIPNKCTPPNVENGILVSDNLSLFSLVNVEFRCPQG 244

QY 223 YEYSEGDVAVCTE-SGWRP-LPSCREE-----KSCDNP 252

Db 245 FVMKGPRRVKCALNKWEPELPCSRVQCQPPDPDLHAERTQDKDNFSPQGEVFYSCEPG 304

QY 253 Y-----IPNGDYSP 261

Db 305 YDLRGAASMRCTPQGDWSP 323

RESULT 15

Q16745

ID Q16745 PRELIMINARY; PRT; 2039 AA.

AC Q16745;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE COMPLEMENT RECEPTOR 1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94065175; PubMed=8245463;

RA Vik D.P., Wong W.W.;

RT "Structure of the gene for the F allele of complement receptor type 1

RT and sequence of the coding region unique to the S allele.";

RL J. Immunol. 151:6214-6224(1993).

RN [2]

RP SEQUENCE FROM N.A.

RX Vik D.P., Wong W.W.;

RA Submitted (JUN-1993) to the EMBL/GenBank/DBDJ databases.

DR EMBL; L17418; AAB60694.1;

DR EMBL; L17390; AAB60694.1; JOINED.

DR EMBL; L17399; AAB60694.1; JOINED.

DR EMBL; L17409; AAB60694.1; JOINED.

DR EMBL; L17419; AAB60694.1; JOINED.

DR EMBL; L17420; AAB60694.1; JOINED.

DR EMBL; L17421; AAB60694.1; JOINED.

DR EMBL; L17422; AAB60694.1; JOINED.

DR EMBL; L17423; AAB60694.1; JOINED.

DR EMBL; L17391; AAB60694.1; JOINED.

DR EMBL; L17392; AAB60694.1; JOINED.

DR EMBL; L17393; AAB60694.1; JOINED.

DR EMBL; L17394; AAB60694.1; JOINED.

DR EMBL; L17395; AAB60694.1; JOINED.

DR EMBL; L17396; AAB60694.1; JOINED.

DR EMBL; L17397; AAB60694.1; JOINED.

DR EMBL; L17398; AAB60694.1; JOINED.

DR EMBL; L17400; AAB60694.1; JOINED.

DR EMBL; L17401; AAB60694.1; JOINED.

DR EMBL; L17402; AAB60694.1; JOINED.

DR EMBL; L17403; AAB60694.1; JOINED.

DR EMBL; L17404; AAB60694.1; JOINED.

DR EMBL; L17405; AAB60694.1; JOINED.

DR EMBL; L17406; AAB60694.1; JOINED.

DR EMBL; L17407; AAB60694.1; JOINED.

DR EMBL; L17408; AAB60694.1; JOINED.

DR EMBL; L17410; AAB60694.1; JOINED.

DR EMBL; L17411; AAB60694.1; JOINED.

DR EMBL; L17412; AAB60694.1; JOINED.

DR EMBL; L17413; AAB60694.1; JOINED.

DR EMBL; L17414; AAB60694.1; JOINED.

DR EMBL; L17415; AAB60694.1; JOINED.

DR EMBL; L17416; AAB60694.1; JOINED.

DR EMBL; L17417; AAB60694.1; JOINED.

DR HSSP; P08603; 1HFI.

DR InterPro; IPR001424; SOD_CU_ZN

DR InterPro; IPR000436; Sushi_SCR_CCP.

DR InterPro; IPR000834; zn-carbopept.

DR Pfam; PF00084; sushi; 30.

DR SMART; SM00032; CCP; 30.

DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.

DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.

KW Receptor.

SQ SEQUENCE 2039 AA; 223603 MW; B82FCB11C6B16635 CRC64;

Query Match 19.9%; Score 298.5; DB 4; Length 2039;

Best Local Similarity 27.0%; Pred. No. 1.4e-19;

Matches 86; Conservative 49; Mismatches 111; Indels 73; Gaps 18;

QY 3 CNE---LPPRRNTEILTGSWSQTYPEGTQAIYKCRPGYRSLGNVTMVCRCGEWALNPL 59

Db 43 CNAPEWLPFARPTNLT- : : | | | : : | | | | | | : : |
QY 60 RCKOKRPGCHPGDTFGTFTLTGGNVFEXGVKAVYTCNEGYYOLLGEINYREC----DTDG 115
Db 98 -RCRRKSCRNPDPVNGVHVYK--IQFGSQIKYISCTKGYRLIGS-SSATCIIISGDTVI 153
QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDREYHFGQAVRVCNSG-----YKIEG 168
Db 154 WNETPICDRIPCGLPPTI-TNGDFISTNRE---NFHYGSVVTYRCNPGSGGRKVFELVG 209
QY 169 DEEMHC--SDD--GFWSKXKPC-VEISCKSPDVINGSPIS-QKIIYKENERFOYKCNMG 222
Db 210 EPSIYCTSNDDQVGIIWSGPAQCIIIPNKCTPPNVENGILVSDNRSLSLNEVVEFEFCQFG 269
QY 223 YEYSERGDAVCTE-SGWRP-LPSCER-----KSCDNP 252
Db 270 FVMKGPRRYKCOALNKWKEPCLPSCSRVCQPPDVLHAERTORDKNFSPQOEVFYSCEPG 329
QY 253 Y-----IPNGDYSP 261
Db 330 YDLRGAAASMRCTPQGDWSP 348

Search completed: August 30, 2002, 06:22:07
Job time: 341 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2002, 06:04:46 ; Search time 29.78 Seconds
(without alignments)
988.400 Million cell updates/sec

Title: US-09-316-163-10

Perfect score: 1497

Sequence: 1 EDNCLPPRRTEILTGWSW.....EKSCDNYPNGDYSPLRIK 265

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802:*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1055	70.5	240	18	AAW39154
2	961	64.2	216	18	AAW39155
3	320	21.4	645	21	AAAB53125
4	312	20.8	263	22	AAAB48846
5	303	20.2	581	12	AAAR13490
6	301.5	20.1	263	20	AAV29859
7	301.5	20.1	263	21	AAAB13014
8	301	20.1	496	20	AAV55752
9	300.5	20.1	543	13	AAAR28557
10	298.5	19.9	543	13	AAAR28547
11	298.5	19.9	543	13	AAAR28567

12	298.5	19.9	778	19	AAW73147	Amino acid sequenc
13	298.5	19.9	1930	19	AAW45899	Human complement r
14	298.5	19.9	2039	20	AAV55751	Human C3b/C4b rece
15	298.5	19.9	2039	22	ABG00287	Novel human diagno
16	298.5	19.9	2044	22	ABB11782	Human CRI protein
17	298.5	19.9	2044	22	AAW39224	Human polypeptide
18	298.5	19.9	2044	22	AAW41010	Human polypeptide
19	298.5	19.9	2317	10	AAV2219	CRI protein. Homo
20	296.5	19.8	543	13	AAAR28560	CRI-4 (114S) analo
21	295.5	19.7	543	13	AAAR28570	CRI-4 (266-274 KLK
22	294.5	19.7	543	13	AAAR28550	CRI-4 (64K) analog
23	294.5	19.7	543	13	AAAR28553	CRI-4 (85R, 87N) a
24	294.5	19.7	543	13	AAAR28565	CRI-4 (121Q) analo
25	294.5	19.7	543	13	AAAR28568	CRI-4 (347T, 349Y)
26	294.5	19.7	543	13	AAAR28569	CRI-4 (369-376 STK
27	294.5	19.7	543	13	AAAR28571	CRI-4 (364-367 NAA
28	294.5	19.7	2039	12	AAAR1810	Human complement t
29	293.5	19.6	543	13	AAAR28545	CRI-4 (37Y) analog
30	293.5	19.6	543	13	AAAR28548	CRI-4 (57V, 59K) a
31	293	19.6	363	18	AAW12414	Porcine complement
32	293	19.6	363	20	AAV30918	MCP protein. Unid
33	292.5	19.5	515	22	AAW93953	Human polypeptide,
34	292.5	19.5	543	13	AAAR28555	CRI-4 (92T) analog
35	292.5	19.5	543	13	AAAR28558	CRI-4 (109N, 110A,
36	292.5	19.5	543	13	AAAR28563	CRI-4 (117P) analo
37	292.5	19.5	1497	22	AAW93954	Human polypeptide,
38	291.5	19.5	543	13	AAAR28549	CRI-4 (64K, 65T) a
39	291.5	19.5	543	13	AAAR28551	CRI-4 (65T) analog
40	291.5	19.5	543	13	AAAR28562	CRI-4 (116K) analo
41	291.5	19.5	543	13	AAAR28566	CRI-4 (318K, 319N)
42	290.5	19.4	254	15	AAAR47154	Sequence of solubl
43	290.5	19.4	254	15	AAAR47155	Deduced sequence o
44	289.5	19.3	263	10	AAV2003	CRI-4 (94H) analog
45	289.5	19.3	543	13	AAAR28556	

ALIGNMENTS

RESULT 1
AAW39154
ID AAW39154 standard; Protein; 240 AA.
XX
AC AAW39154;
XX
DT 27-APR-1998 (first entry)
XX
DE Human partial Complement factor H protein fragment 1.
XX
DE Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medicament; modulator.
XX
OS Homo sapiens.
XX
PN WO9738136-A1.
XX
PD 16-OCT-1997.
XX
PF 09-APR-1997; 97WO-US05710.
XX
PR 06-MAR-1997; 97US-0812481.
PR 09-APR-1996; 96US-0015083.
PR 09-APR-1996; 96US-0630048.
PR 06-MAR-1997; 97US-0038614.
XX
PA (BARD-) BARD DIAGNOSTIC SCI INC.
XX
XX Enfield DL, Hass GM, Kinders RJ;
PI WPI: 1997-512742/47.
DR N-PSDB; AAV02790.
XX
PT Treating or screening for cancer, e.g. renal or urogenital cancer -

PT by modulating or detecting tumour associated human complement Factor
PT H related antigen, or nucleic acid encoding it
PS
PS Example 6B; Fig 6B; 104pp; English.
XX
XX This partial protein sequence represents a region of the human
CC tumour-associated complement factor H (CFH). This sequence is used
CC in the identification of complement factor H related proteins and
CC antigens isolated from clone PRBB9FH410 (see AAW39155). The detection of
CC such proteins and a CFH antigens can be used in screening or for the
CC treatment of renal or urogenital cancer, e.g. bladder, cervical or
CC prostate cancer. Agents that may modulate this antigen could be used in
CC the manufacture of a medicament for the treatment of a tumour cell.
XX
XX Sequence 240 AA;
SQ

Query Match 70.5%; Score 1055; DB 18; Length 240;
Best Local Similarity 100.0%; Pred. No. 1e-80;
Matches 188; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 FTLTGGNVFYGKAVYTCNEGKOLLGEINRYRECDTGDWTDIPICEVVKCLPVTAPENG 137
DB 1 fcltggnvfygkavycnegyqlgeinryrecdtgdwtndipicevkvclpvtapeng 60

QY 138 KIVSSAMEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFWSKPKVCVEISCKSPDV 197
DB 61 kivssamepdrehyhfgqavrfvcnsgyklegdeemhcsddgfwskpkvcveisckspdv 120

QY 198 INGSPTSOIKIYKENERFOYKCNMGVEYSERGDVCTESGWRPLPSCEEKSCDNPYIPNG 257
DB 121 inspsisqkilykenerfkykcnmgveysersergdavctesgwrplpsceekscdnpyipng 180

QY 258 DYSPLRIK 265
DB 181 dysplrik 188

RESULT 2
AAW39155
ID AAW39155 standard; Protein; 216 AA.
XX
XX AAW39155;
XX
DT 27-APR-1998 (first entry)
DE Clone PRBB9FH410 CFH related protein fragment.
KW Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medicament; modulator.
XX
XX Synthetic.
XX
XX WO9738136-A1.
XX
PD 16-OCT-1997.
XX
PF 09-APR-1997; 97WO-US05710.
XX
PR 06-MAR-1997; 97US-0812481.
PR 09-APR-1996; 96US-0015083.
PR 09-APR-1996; 96US-0630048.
PR 06-MAR-1997; 97US-0038614.
XX
PA (BARD-) BARD DIAGNOSTIC SCI INC.
XX
PI Enfield DL, Hass GM, Kinders RJ;
XX
DR WPI; 1997-512742/47.
DR N-PSDB; AAW02791.
XX
XX Treating or screening for cancer, e.g. renal or urogenital cancer -
PT by modulating or detecting tumour associated human complement Factor

PT H related antigen, or nucleic acid encoding it
XX
PS Example 6B; Fig 6B; 104pp; English.
XX
XX This partial protein is found in clone PRBB9FH410 and represents a
CC complement factor H related protein with homology to a region of the
CC human tumour-associated complement factor H (CFH). The detection of this
CC protein and a CFH antigen can be used in screening or for the treatment
CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.
CC Agents that may modulate this antigen could be used in the manufacture of
CC a medicament for the treatment of a tumour cell.
XX
XX Sequence 216 AA;
SQ

Query Match 64.2%; Score 961; DB 18; Length 216;
Best Local Similarity 99.4%; Pred. No. 7e-73;
Matches 170; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 95 TCNEGKOLLGEINRYRECDTGDWTDIPICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQ 154
DB 1 tcnegyqlgeinryrecdtgdwtndipicevkvclpvtapengkivssamepdreynf9q 60

QY 155 AVRFVCSNGYKIEGDEEMHCSDDGFWSKPKVCVEISCKSPDVINGSPISQKIIYKENER 214
DB 61 avrfvcnsgyklegdeemhcsddgfwgkpkvcveisckspdvingspsqkiiykener 120

QY 215 FOYKCNMGVEYSERGDVCTESGWRPLPSCEEKSCDNPYIPNGDYSPLRIK 265
DB 121 fgykcnmgveysersergdavctesgwrplpsceekscdnpypngdysplrik 171

RESULT 3
AAB53125
ID AAB53125 standard; Protein; 645 AA.
XX
XX AAB53125;
XX
DT 28-FEB-2001 (first entry)
DE Macaca mulatta rhadinovirus 17577 RRV ORF4 protein SEQ ID NO:7.
XX
KW Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;
KW genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;
KW IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;
KW cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;
KW lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
KW splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia.
XX
XX Macaca mulatta rhadinovirus 17577.
XX
XX WO200028040-A2.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26260.
XX
PR 06-NOV-1998; 98US-0107507.
PR 20-NOV-1998; 98US-0109409.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Wong SW, Axthelm MK, Searles RP;
XX
XX WPI; 2000-376552/32.
XX
XX New rhesus rhadino virus for producing non-human primate model useful
PT for testing potential treatments and efficacy of the candidate vaccine
PT for conditions associated with RRV infection -
XX
XX Claim 5; Page 122-123; 141pp; English.
XX
XX The present invention describes a novel rhesus macaque rhadinovirus

CC additive to blood, e.g. in an extracorporeal circulation system (coated
CC on tubing) or in storage, also for studying complement activation.
CC Transgenic animals that express SPICE are used as sources of xenografts.
CC The present sequence represents a vaccinia complement control protein
CC (VCP) encoded by the specifically claimed mutated VCP nucleotide
CC sequence, having a silent T to A transversion at nucleotide position
CC number 267.
XX
SQ Sequence 263 AA;

Query Match 20.1%; Score 301.5; DB 20; Length 263;
Best Local Similarity 33.3%; Pred. No. 1.9e-17;
Matches 78; Conservative 29; Mismatches 106; Indels 21; Gaps 12;

QY 20 SDQTYPEGTQAIYKCRPGYR--SLGNVIMVCRKGEWALNPLRKCKOKRPGDTPFGT 77
DB 40 ananynigdtieylclpgyrkqkmpiyaktgtwllfn---qcikrrcpsprldnng 96
QY 78 FTLTGNNVFEYGVKAVYTCNEGQQLLGEI-NYRECDTDG---WTNDIPICEVVKCLPVT 133
DB 97 lldigg---vdfgssitcyscnsghlligesksycelgstgsmvwnpeapicesvkcsqps 153
QY 134 PENGKIVSSAMEPDRYHFQAVRFVCNCGYKIEGDEEMHCSDDGFWSEKPKCIVEISCK 193
DB 154 isngr--hngyed--fytdgsvvtyscnsghlligesksycelgstgsmvwnpeapicesvkcsqps 207
QY 194 SPDVINGSPISQ-KIIYKENERFOYKCNMGYSEYSESGDAVCTESG-WRP-LPSC 244
DB 208 hptisngylssgfkrsysyndvdfckgykylsgssstcspgntwkpkelpkc 261

RESULT 7
AAB13014
ID AAB13014 standard; protein; 263 AA.
XX
AC AAB13014;
XX
DT 11-DEC-2000 (first entry)
XX

DE Complement inhibitory protein VCP amino acid sequence.

KW Alzheimer's disease; Vaccinia virus; VCP; complement pathway inhibitor;
KW treatment; diagnosis; amyloid plaque.

XX Vaccinia virus.

XX WO2000043027-A1.

XX 27-JUL-2000.

XX 19-JAN-2000; 2000WO-US01115.

XX 19-JAN-1999; 99US-0116328.

XX (UYLO-) UNIV LOUISVILLE RES FOUND INC.

PA (KOTW/) KOTWAL G J.

PA (DALY/) DALY J.

XX Kotwal GJ, Daly J;

XX WPI; 2000-476187/41.

DR N-PSDB; AAA72738.

XX Treating Alzheimer's disease using a Vaccinia virus protein that blocks

PT the complementation pathway

XX Claim 1; Page 42-43; 96pp; English.

XX This invention relates to a method for treating Alzheimer's disease. The
CC method uses a composition comprising a Vaccinia virus complement control
CC protein, which blocks the complement pathway by binding to complement
CC components. The protein designated VCP, blocks complement activation and

CC can bind to amyloid plaques in samples. The transmembrane domain and a
CC portion of the C-terminus (Abeta) of the amyloid precursor protein (APP)
CC form the nucleus of the amyloid plaque, the hallmark of Alzheimer's
CC disease. The beta peptide activates the complement pathway. The VCP
CC protein used in the method of the invention, down regulates the
CC complement pathway activation caused by Abeta. The method and composition
CC may be used for the treatment of Alzheimer's disease. The VCP protein
CC also binds to amyloid plaques and may be labelled and used to detect the
CC presence of amyloid plaques in pathological samples, and therefore
CC diagnose Alzheimer's disease. The present sequence represents the
CC Vaccinia virus VCP protein used in the method of the invention.
XX
SQ Sequence 263 AA;

Query Match 20.1%; Score 301.5; DB 21; Length 263;
Best Local Similarity 33.3%; Pred. No. 1.9e-17;
Matches 78; Conservative 29; Mismatches 106; Indels 21; Gaps 12;

QY 20 SDQTYPEGTQAIYKCRPGYR--SLGNVIMVCRKGEWALNPLRKCKOKRPGDTPFGT 77
DB 40 ananynigdtieylclpgyrkqkmpiyaktgtwllfn---qcikrrcpsprldnng 96
QY 78 FTLTGNNVFEYGVKAVYTCNEGQQLLGEI-NYRECDTDG---WTNDIPICEVVKCLPVT 133
DB 97 lldigg---vdfgssitcyscnsghlligesksycelgstgsmvwnpeapicesvkcsqps 153
QY 134 PENGKIVSSAMEPDRYHFQAVRFVCNCGYKIEGDEEMHCSDDGFWSEKPKCIVEISCK 193
DB 154 isngr--hngyed--fytdgsvvtyscnsghlligesksycelgstgsmvwnpeapicesvkcsqps 207
QY 194 SPDVINGSPISQ-KIIYKENERFOYKCNMGYSEYSESGDAVCTESG-WRP-LPSC 244
DB 208 hptisngylssgfkrsysyndvdfckgykylsgssstcspgntwkpkelpkc 261

RESULT 8
AAY5752
ID AAY5752 standard; Protein; 496 AA.
XX
AC AAY5752;
XX
DT 22-FEB-2000 (first entry)
XX

DE Human CR1 protein LHR-A SCR fragment.

KW C3B/C4B receptor; CR1 protein; cell-surface protein; erythrocyte; human;
KW complement regulatory activity; complement pathway enzyme; tissue damage;
KW reperfusion injury; Arthus reaction; myocardial infarct; inflammation;
KW heart condition; autoimmune disorder; long homologous repeat; LHR; SCR;
KW short consensus repeat.

XX Homo sapiens.

XX US5981481-A.

XX 09-NOV-1999.

XX 06-JUN-1995; 95US-0470652.

XX 03-APR-1989; 89US-0332865.

PR 06-DEC-1974; 74US-0350238.

PR 24-FEB-1993; 93US-0026134.

PR 01-APR-1988; 88US-0176532.

XX (UYJO) UNIV JOHNS HOPKINS.

PA (BGHM) BRIGHAM & WOMENS HOSPITAL.

PA (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.

XX Concino MF, Wong WM, Makrides SC, Klickstein LB, Fearon DT, Ip SH;

PI Marsh HC, Carson GR;

XX WPI; 1999-633357/54.

Db 289 ydlrgaasmrctpgdwsapaaptecvksdcd 319

Matches 88; Conservative 46; Mismatches 102; Indels 95; Gaps 18;

QY 3 CNE---LPPRRNTTEILTGSMSDQTYPEGTQAIYKRCGRSLGNVIVCRKGEVVALNPL 59
 Db 2 cnapewlpfarptnit-----defefpigtynyecrpysgrpfslclksv--sspk 55
 QY 60 RKQKRPCGHPGDPFGTFTLTGNGVFEYGVKAVYTCNEGQOLLGEINRYEC-----DTDG 115
 Db 56 drcrkscrnpdpvngmvhvkq--lqfgsqikysctkgryllgs-ssatcillsgdv1 112
 QY 116 WTNDIPICEVVVKC-LPVTAPENGKIVSSAMEPDRYHFGQAVRFVNCNG-----YKIEG 168
 Db 113 wdnetspicdrpcglpbt1-tngdfistnre---nfhygsvvtvrcnpgsggrkvfelvg 168
 QY 169 DEEMHC--SDD--GFWSKEXPKC-VEISCKSPDVINGSPIS-OKIIVKEN----- 212
 Db 169 epsiycstnddqvglwsgpapqclpnkctppnvengllvsdnrlsflsnevefrqcpv 228
 QY 213 -----ERFQ-----YKCNMG 222
 Db 229 fvmkgprvrkcaqlnkweplscsvrcvqppdv1haertqrkdntfsggqevfyscepg 288
 QY 223 YEYSERGDAVCTESG-WRP-LPSCEEKSCDN 251
 Db 289 ydlrgaasmrctpgdwsapaaptecvksdcd 319

RESULT 11

AAR28567

ID AAR28567 standard; peptide; 543 AA.

XX

AC AAR28567;

XX

DT 19-MAR-1993 (first entry)

XX

DE CRI-4 (318-321 RNPP) analogue.

XX

KW short consensus repeat; regulator of complement activation;

KW C3b binding; C4b binding; human complement type 1 receptor.

XX

OS Homo sapiens.

XX

PH Key Location/Qualifiers

FT Region 1..60

FT T /label= SCR-1

FT Region 61..122

FT T /label= SCR-2

FT Region 451..510

FT T /label= SCR-8

FT Region 511..543

FT T /label= SCR-9

FT T /note= "TRUNCATED"

FT Misc-difference 318

FT T /note= "Asp substituted by Arg"

FT Misc-difference 319

FT T /note= "Asp substituted by Asn"

FT Misc-difference 320

FT T /note= "Phe substituted by Pro"

FT Misc-difference 321

FT T /note= "Met substituted by Pro"

XX

PN EP512733-A.

XX

PD 11-NOV-1992.

XX

PF 28-APR-1992; 92EP-0303826.

XX

PR 03-MAY-1991; 91US-0695514.

XX

PA (UNIW) UNIV WASHINGTON.

XX

PI Atkinson JP, Hourcade D, Krych M;

XX

DR WPI; 1992-375009/46.

XX

DT Complement activity regulator protein analogues - useful for treating auto-immune diseases, to suppress transplant rejection, for diagnosis etc.

XX

PS Claim 11; Fig 2 and R11810; 23pp; English.

XX

CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988) 168:1255-1270. It encodes the first 8 and a half amino terminal SCRs of CRI. The invention concerns analogues of "regulator of complement activation" proteins or truncated, hybrid or recombinant forms of them. CRI-4 is a preferred truncated form and a number of specified substitution variants of it are claimed in which certain positions in SCR-1 which have been identified as important for the degree of C3b- and C4b-binding are substituted by amino acids from the corresponding positions in SCR-8. The specification does not contain the CRI-4 sequence; the sequence given here was constructed from the full-length CRI amino acid sequence having GENESEQ accession number AAR11810 and descriptions in the disclosure.

XX

CC Sequence 543 AA;

SQ

Query Match 19.9%; Score 298.5; DB 13; Length 543;

Best Local Similarity 26.6%; Pred. No. 8.5e-17;

```

DR  WPI; 1992-375009/46.
XX
PT  Complement activity regulator protein analogues - useful for
PT  treating auto-immune diseases, to suppress transplant rejection,
XX  for diagnosis etc.
XX
PS  Example 8; Page 18 and R11810; 23pp; English.
XX
CC  The cDNA clone designated CR1-4 was described in J.Exp.Med.(1988)
CC  168:1255-1270. It encodes the first 8 and a half amino terminal
CC  SCR's of CR1. The invention concerns analogues of "regulator of
CC  complement activation" proteins or truncated, hybrid or recombinant
CC  forms of them. CR1-4 is a preferred truncated form and a number of
CC  specified substitution variants of it are disclosed in which certain
CC  positions in SCR-5-6 are substituted by amino acids from
CC  the corresponding positions in SCR's which are involved in C3b- and
CC  C4b-binding. The substitution variant given here has increased C4b-
CC  binding. The substitution does not contain the CR1-4 sequence;
CC  the sequence given here was constructed from the full-length CR1
CC  amino acid sequence having GENESQ accession number AAR11810 and
CC  descriptions in the disclosure.
XX
SQ  Sequence 543 AA;

Query Match      19.9%; Score 298.5; DB 13; Length 543;
Best Local Similarity 26.5%; Pred. No. 8.5e-17;
Matches 88; Conservative 44; Mismatches 105; Indels 95; Gaps 18;

QY  3 CNE---LPPRNTTEILTGSWSQDTYPECTQAIYKRCPCYRSLGNVIMVCRKGWALNPL 59
DB  || || || || || || || || || || || || || || || || || || || || ||
DB  2 cnapewlpfartnlt-----defefpigtynyecrpysgrpfaiiclknsvtgskd- 56
QY  60 RKQKRCPCGHGPDGPFPGFTTLTGGNVFYGKAVTTCNEGYYQLLGEINRYREC----DTDG 115
DB  || || || || || || || || || || || || || || || || || || || || ||
DB  57 -rcrrksrnpdpvngmvhvikg--iqfsgqikysctkgyrligs-ssatcilsgdvti 112
QY  116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDREYHFGQAVRFVCSNG-----YKIEG 168
DB  || || || || || || || || || || || || || || || || || || || || ||
DB  113 wdnetpicdrpcglpplti-tngdfistnre---nfhygsvtvtrcnpdgsgrkvfelvg 168
QY  169 DEEMHC--SDD--GFWSKEKPKC-VEISCKSPDIVNGSPIS-QKIYKEN-----212
DB  || || || || || || || || || || || || || || || || || || || || ||
DB  169 epslyclnddqvlgwsgpqpclpnkctppnvengilvdsnrslfslnevvfrcqpv 228
QY  213 -----ERFQ-----YKCNMG 222
DB  || || || || || || || || || || || || || || || || || || || || ||
DB  229 fvmkgprvrkcaqlnkwepeipscsvrcvqppdvhlhaertgrdkdnfsgqevfyscepg 288
QY  223 YEYSEGDVCTESG-WRP-LPSCSEKSCDNP 252
DB  || || || || || || || || || || || || || || || || || || || || ||
DB  289 ydlrgaasmrctpgdgwspapaptcevkscrnp 320

RESULT 12
AAW73147
ID  AAW73147 standard; protein; 778 AA.
XX
AC  AAW73147;

XX  29-JAN-1999 (first entry)
XX
DE  Amino acid sequence of the soluble complement receptor 1 (sCR1).
XX
KW  Human; soluble complement receptor 1; sCR1; T-cell; B-cell;
KW  mediated immune response; inhibition; tissue rejection; gene therapy;
KW  dystrophin; inflammatory response; interferon-gamma secretory response;
KW  autoimmune response; neurological response; Alzheimer's disease;
KW  Parkinson's disease; multiple sclerosis; systemic lupus erythematosus;
KW  rheumatoid arthritis; myasthenia gravis; epidermis bullosa;
XX  Hashimoto's disease.
XX
OS  Homo sapiens.

```

```

XX  W09845430-A1.
XX
XX  15-OCT-1998.
XX
XX  06-APR-1998; 98WO-GB01012.
XX
XX  05-APR-1997; 97GB-0006950.
XX
XX  (ANNE/) ANNENKOV A.
XX  (CHER/) CHERNAJOVSKY Y.
XX
XX  Annenkov A, Chernajovsky Y;
XX  WPI; 1998-568350/48.
XX
XX  Fragment of soluble human complement receptor 1 - useful for
XX  treating T-cell or B-cell mediated immune responses e.g.
XX  inflammatory responses such as rheumatoid arthritis
XX
XX  Disclosure; Fig 1; 54pp; English.
XX
XX  This is an amino acid sequence of the human soluble complement
XX  receptor 1 (sCR1), useful in the treatment of T-cell or B-cell
XX  mediated immune responses. It is used to inhibit a T-cell or
XX  B-cell-mediated immune response to prevent immune response-mediated
XX  tissue rejection and destruction or clearance or inactivation of an
XX  expressed protein especially from cells that have been treated by an
XX  therapy to express the protein, e.g. dystrophin. The protein can also
XX  be used to inhibit a T-cell or B-cell-mediated inflammatory response,
XX  an interferon-gamma secretory response, autoimmune response or
XX  neurological response, e.g. Alzheimer's or Parkinson's disease or
XX  multiple sclerosis. Also the protein can be used to treat systemic
XX  lupus erythematosus, rheumatoid arthritis, myasthenia gravis,
XX  epidermis bullosa or Hashimoto's disease.
XX
SQ  Sequence 778 AA;

Query Match      19.9%; Score 298.5; DB 19; Length 778;
Best Local Similarity 27.0%; Pred. No. 1.3e-16;
Matches 86; Conservative 49; Mismatches 111; Indels 73; Gaps 18;

QY  3 CNE---LPPRNTTEILTGSWSQDTYPECTQAIYKRCPCYRSLGNVIMVCRKGWALNPL 59
DB  || || || || || || || || || || || || || || || || || || || || ||
DB  48 cnapewlpfartnlt-----defefpigtynyecrpysgrpfaiiclknsvtgskd- 102
QY  60 RKQKRCPCGHGPDGPFPGFTTLTGGNVFYGKAVTTCNEGYYQLLGEINRYREC----DTDG 115
DB  || || || || || || || || || || || || || || || || || || || || ||
DB  103 -rcrrksrnpdpvngmvhvikg--iqfsgqikysctkgyrligs-ssatcilsgdvti 158
QY  116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDREYHFGQAVRFVCSNG-----YKIEG 168
DB  || || || || || || || || || || || || || || || || || || || || ||
DB  159 wdnetpicdrpcglpplti-tngdfistnre---nfhygsvtvtrcnpdgsgrkvfelvg 214
QY  169 DEEMHC--SDD--GFWSKEKPKC-VEISCKSPDIVNGSPIS-QKIYKENRFQYKCNMG 222
DB  || || || || || || || || || || || || || || || || || || || || ||
DB  215 epslyclnddqvlgwsgpqpclpnkctppnvengilvdsnrslfslnevvfrcqpv 274
QY  223 YEYSEGDVCTE-SGWRP-LPSCSE-----KSCDNP 252
DB  || || || || || || || || || || || || || || || || || || || || ||
DB  275 fvmkgprvrkcaqlnkwepeipscsvrcvqppdvhlhaertgrdkdnfsgqevfyscepg 334
QY  253 Y-----IPNGDYSP 261
DB  || || || || || || || || || || || || || || || || || || || || ||
DB  335 ydlrgaasmrctpgdgwsp 353

RESULT 13
AAW45899
ID  AAW45899 standard; peptide; 1930 AA.
XX
AC  AAW45899;

```


Query Match	19.98;	Score 298.5;	DB 20;	Length 2039;
Best Local Similarity	27.08;	pred. No. 4.3e-16;		
Matches	86;	Conservative 49;	Mismatches 111;	Indels 73; Gaps 18;
Qy	3	CNE---LPPRNTTEILTGSWSDDQTYPECTQAIYKCRDQYRSLGNVIMVCRKGEWVALNPL	59	
Db	43	cnapewlpfaaptntl---defeflgtlylnyecrpgysgrpfslclcnksvvtgakd-	97	
Qy	60	RKCKRRCPGHPGDTPFTLTGTGNVFEYGVKAVTTCNEGYYQLLGEINVRBC---	115	
Db	98	-rrrrckrcnppdvngmvhvik--lqfsgikysetckgyrlgs--ssatciisgdtvi	153	
Qy	116	WTNDIPICEVVKC-LPVTAPNGKIVSAMEDREYHFGQAVRVVCNSG-----	168	
Db	154	wdnetpicdrilpcglppti-tngdfisltnre---nfnysvvtvrcnpgsggrkvfeivg	209	
Qy	169	DEEMHC--SDD--GFWSKEPKC-VEISCKSPDVINGSPI-SQIIYKENERFOYKCNMG	222	
Db	210	epslyctsnddqvgiawsgppqclpnkctppnvengillvsdnrlslfslnevvfrcqpg	269	
Qy	223	YEYSERGDAVCTE-SGNRP-LPSCSE-----KSCDNP	252	
Db	270	fvmkgprvrvcqalnkwepeipscsrvcqppdvlnhaertqrdknfsgqevfyscepg	329	
Qy	253	Y-----IPNGDYSP	261	
Db	330	ydlrgaasmrctpgqdwsp	348	

RESULT	15
ABG00287	
ID	ABG00287 standard; Protein; 2039 AA.
XX	
AC	ABG00287;
XX	
DT	13-FEB-2002 (first entry)
XX	
DE	Novel human diagnostic protein #278.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
FD	11-OCT-2001.
XX	
XX	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drmanac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
DR	N-PSDB; AAS64474.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	
PS	Claim 20; SEQ ID NO 30646; 103pp; English.

to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Search completed: August 30, 2002, 06:18:45
Job time: 839 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:02:15 ; Search time 39.15 Seconds
(without alignments)
807.494 Million cell updates/sec

Title: US-09-316-163-11

Perfect score: 1876
Sequence: 1 EDNCWELPPRRNTTEILTGSWS.....PDIKHGGLYHENMRPYPVPV 329

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1876	100.0	449	1 NBHUS	complement factor
2	1876	100.0	1231	1 NBHUS	complement factor
3	1261	67.2	1234	1 NBMS	complement factor
4	903	48.1	669	2 S6551	factor H - bovine
5	486	25.9	1053	2 S46199	probable complemen
6	356	19.0	597	1 S53711	C4BP alpha chain p
7	353	18.8	597	1 NBHUC4	C4b-binding protei
8	353	18.8	676	2 A45900	complement C3b rec
9	350.5	18.7	2489	2 I73012	complement C3b/c4b
10	348	18.6	2014	2 I36936	complement recepto
11	343.5	18.3	482	2 A34924	complement C3b/c4b
12	337.5	18.0	560	2 T16833	hypothetical prote
13	335	17.9	360	2 T42921	complement control
14	323	17.2	497	2 J62054	complement regulat
15	320.5	17.1	610	1 I46001	C4b-binding protei
16	317	16.9	452	2 A35068	complement factor
17	317	16.9	558	2 S57953	C4BP protein alpha
18	311.5	16.6	469	1 NBMSCA	C4b-binding protei
19	306	16.3	303	2 H35068	apolipoprotein H-r
20	304	16.2	440	2 A43519	complement recepto
21	303.5	16.2	1025	1 A43526	complement C3d/Eps
22	303	16.2	661	1 KFHUI3	apolipoprotein H h
23	301.5	16.1	263	1 WMVZSP	coagulation factor
24	294	15.7	668	2 A46013	coagulation factor
25	288	15.4	1091	1 PL0009	complement C3d/Eps
26	286	15.2	830	2 A30359	P-selectin precurs
27	285	15.2	579	2 A56740	sperm-egg recognit
28	284.5	15.2	263	1 C36838	complement control
29	284.5	15.2	263	2 T28450	hypothetical prote

30	283.5	15.1	263	2 B72152	B18L protein - var
31	279.5	14.9	345	1 NBMS	apolipoprotein H p
32	279	14.9	610	2 A35046	E-selectin precurs
33	276.5	14.7	612	2 B42755	E-selectin precurs
34	274.5	14.6	349	2 G02913	sperm CD46 - human
35	274.5	14.6	369	2 I57998	membrane cofactor
36	274.5	14.6	768	2 A42755	membrane cofactor
37	274	14.6	377	2 I54479	membrane cofactor
38	274	14.6	384	2 S01896	P-selectin - rat
39	274	14.6	768	2 I53821	membrane cofactor
40	270.5	14.4	362	2 JC5194	membrane cofactor
41	270.5	14.4	369	2 JC5138	membrane-bound com
42	269.5	14.4	360	1 WMBE2E	secretory compleme
43	267.5	14.3	302	1 WMBE1E	apolipoprotein H p
44	267.5	14.3	345	1 NBHU	E-selectin - pig
45	266	14.2	482	2 JC5092	

ALIGNMENTS

RESULT 1

NBHUS

complement factor H precursor, short splice form [validated] - human
N:Alternate names: Complement factor H-related protein; complement protein H
C:Species: Homo sapiens (man)
C>Date: 31-Dec-1993 #sequence_revision 23-Feb-1996 #text_change 08-Dec-2000
C:Accession: S03013; B60238; A27877; A61103; A26505; S10479
R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.

Biochem. J. 249, 593-602, 1988

A:Title: The complete amino acid sequence of human complement factor H.

A:Reference number: S00254; MUID:88134059

A:Accession: S03013

A:Molecule type: mRNA

A:Residues: 1-449 <RIP>

A:Cross-references: EMBL:X07523; EMBL:Y00716; NID:g32492; PIDN:CAA30403.1; PID:g75807

A:Note: part of this sequence, including the amino end of the mature protein was conf

A:Note: 402-Tyr was also found

R:Estaller, C.; Schwaible, W.; Dierich, M.; Weiss, E.H.

Eur. J. Immunol. 21, 799-802, 1991

A:Title: Human complement factor H: two factor H proteins are derived from alternativ

A:Reference number: A60238; MUID:91184292

A:Accession: B60238

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-33;434-449 <EST>

A:Note: only portions of this 1.8 kilobase mRNA were sequenced

R:Schulz, T.F.; Schwaible, W.; Stanley, K.K.; Weiss, E.; Dierich, M.P.

Eur. J. Immunol. 16, 1351-1355, 1986

A:Title: Human complement factor H: isolation of cDNA clones and partial cDNA sequenc

A:Reference number: A27877; MUID:87054207

A:Accession: A27877

A:Molecule type: mRNA

A:Residues: 1-55;401-'V',403-449 <SCH>

A:Cross-references: GB:X04697; NID:g31991; PIDN:CAB41739.1; PID:g4725976

A:Note: an additional nucleotide present within the codon for Glu-310 was thought to

R:Schwaible, W.; Zwirner, J.; Schulz, T.F.; Linke, R.P.; Dierich, M.P.; Weiss, E.H.

Eur. J. Immunol. 17, 1485-1489, 1987

A:Title: Human complement factor H: expression of an additional truncated gene produc

A:Reference number: A26505; MUID:88055295

A:Accession: A61103

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 27-76 <SC2>

A:Note: this is a partial sequence of an alternatively spliced 1.8 kilobase mRNA that

R:Sim, R.B.; Discipio, R.G.

Biochem. J. 205, 285-293, 1982

A:Title: Purification and structural studies on the complement-system control protein

A:Reference number: A26505; MUID:83048213

A:Accession: A26505

A:Molecule type: protein

A:Residues: 19-20,'Q',22-29,'V',31-33,'Q',35 <SIM>

R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P

Biochemistry 31, 3626-3634, 1992
A:Title: Solution structure of the fifth repeat of factor H: A second example of the complement factor H precursor, long splice form [validated] - human
A:Reference number: A44551; MUID:92232649
A:Contents: annotation: NMR structure determination, residues 264-292
R:Kristensen, T.; Wetsel, R.A.; Tack, B.F.
J. Immunol. 136, 3407-3411, 1986
A:Title: Structural analysis of human complement protein H: homology with C4b binding protein
A:Reference number: S10479; MUID:86169701
A:Accession: S10479
A:Molecule type: mRNA
A:Residues: 226-401, 'Y', 403-449 <KRI>
A:Cross-references: GB:M12383; NID:gl80472; PIDN:AAA52013.1; PID:gl80473
A:Comment: Factor H has also been found bound to cell membranes in an unknown manner. Homologous proteins have also been found in liver, kidney, and placenta.
C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver.
C:Genetics: <HF1>
A:Gene: GDB:HF1; HF
A:Map position: lq32-lq32
A:Cross-references: GDB:120041; OMIM:134370
C:Genetics: <HF2>
A:Gene: GDB:HF2; HF
A:Cross-references: GDB:129095
A:Map position: lq32-lq32
A:Note: the correspondence between the two loci and the sequences indicated is unclear; the alternative complement pathway
C:Function:
A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also increases the alternative complement pathway
A:Pathway: complement alternate pathway
C:Superfamily: complement factor H; complement factor H repeat homology
C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-449/Product: complement factor H, short splice form #status experimental <MAT>
F:21-80/Domain: complement factor H repeat homology <FH01>
F:85-141/Domain: complement factor H repeat homology <FH02>
F:146-205/Domain: complement factor H repeat homology <FH03>
F:210-262/Domain: complement factor H repeat homology <FH04>
F:246-248/Region: cell attachment (R-G-D) motif
F:267-320/Domain: complement factor H repeat homology <FH05>
F:325-385/Domain: complement factor H repeat homology <FH06>
F:389-442/Domain: complement factor H repeat homology <FH07>
F:21-66,52-80,85-129,114-146,192,178-205,210-251,237-262,267-309,294-320,325-374,357
F:217/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 100.0%; Score 1876; DB 1; Length 449;
Best Local Similarity 100.0%; Pred. No. 1.1e-126;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EDCNELPPRRNTTEILGSGNSDQTPYEGTQAIYKCRPGYSLGNVIMVCRKGWVALNPLR 60
DB 19 EDCNELPPRRNTTEILGSGNSDQTPYEGTQAIYKCRPGYSLGNVIMVCRKGWVALNPLR 78
QY 61 KCOKRCGHPGDPFTFTLTGNGVPEYGVKAVYTCNEGYYQLLGEINRECDTGDGTNDI 120
DB 79 KCOKRCGHPGDPFTFTLTGNGVPEYGVKAVYTCNEGYYQLLGEINRECDTGDGTNDI 138
QY 121 PICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVCSNGYKLEGDEMHCSDDGFW 180
DB 139 PICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVCSNGYKLEGDEMHCSDDGFW 198
QY 181 SKEKPKCVESCKSPDVINGSPISQIIYKENERFOYKCMGYEYSGRDAVCTESGWRP 240
DB 199 SKEKPKCVESCKSPDVINGSPISQIIYKENERFOYKCMGYEYSGRDAVCTESGWRP 258
QY 241 LPSCEKSCDNPYPINGDYSPLRKIRTKTGDEITYQCRNGFYPATRNTAKCTSTGWIPAP 300
DB 259 LPSCEKSCDNPYPINGDYSPLRKIRTKTGDEITYQCRNGFYPATRNTAKCTSTGWIPAP 318
QY 301 RCTLPCDYPDKHGLYHNNRRPFPV 329
DB 319 RCTLPCDYPDKHGLYHNNRRPFPV 347

RESULT 2

NBHUH
Complement factor H precursor, long splice form [validated] - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence-revision 31-Dec-1993 #text-change 08-Dec-2000
A:Reference number: A44551; MUID:92232649
A:Contents: annotation: NMR structure determination, residues 264-292
R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.
Biochem. J. 249, 593-602, 1988
A:Title: The complete amino acid sequence of human complement factor H.
A:Reference number: S00254; MUID:88134059
A:Accession: S00254
A:Molecule type: mRNA
A:Residues: 1-1231 <RIP>
A:Cross-references: EMBL:Y00716; NID:g31964; PIDN:CAA68704.1; PID:g31965
A:Note: 402-Tyr was also found
A:Note: parts of this sequence, including the amino and carboxyl ends of the mature
R:Estaller, C.; Schwaible, W.; Dierich, M.; Weiss, E.H.
Eur. J. Immunol. 21, 799-802, 1991
A:Title: Human complement factor H: two factor H proteins are derived from alternative
A:Reference number: A60238; MUID:91184292
A:Accession: A60238
A:Molecule type: mRNA
A:Residues: 1-56;1177-1231 <EST>
A:Note: only portions of this 4.3 kilobase mRNA were sequenced
R:Day, A.J.; Ripoche, J.; Lyons, A.; McIntosh, B.; Harris, T.J.R.; Sim, R.B.
Biosci. Rep. 7, 201-207, 1987
A:Title: Sequence analysis of a cDNA clone encoding the C-terminal end of human comp
A:Reference number: A54726; MUID:88025472
A:Accession: A54726
A:Molecule type: mRNA
A:Status: not compared with conceptual translation
A:Residues: 'DFRN', 579-1231 <DAY>
A:Cross-references: GB:M17517; NID:gl80497; PIDN:AAA52016.1; PID:gl80498
A:Note: parts of this sequence were determined by protein sequencing
R:Ripoche, J.; Day, A.J.; Willis, A.C.; Belt, K.T.; Campbell, R.D.; Sim, R.B.
Biosci. Rep. 6, 65-72, 1986
A:Title: Partial characterization of human complement factor H by protein and cDNA s
A:Reference number: A61565; MUID:86188123
A:Accession: A61565
A:Molecule type: mRNA
A:Status: not compared with conceptual translation
A:Residues: 'METGRNHLNAKI', 1050-1057, 'T', 1059-1102 <RI2>
R:Sim, R.B.; Discipio, R.G.
Biochem. J. 205, 285-293, 1982
A:Title: Purification and structural studies on the complement-system control protei
A:Reference number: A26505; MUID:83048213
A:Accession: A26505
A:Molecule type: protein
A:Residues: 19-20, 'Q', 22-29, 'V', 31-33, 'Q', 35 <SIM>
R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll,
Biochemistry 31, 3626-3634, 1992
A:Title: Solution structure of the fifth repeat of factor H: A second example of the
A:Reference number: A44551; MUID:92232649
A:Contents: annotation: NMR structure determination, residues 264-292
R:Norman, D.G.; Barlow, P.N.; Baron, M.; Day, A.J.; Sim, R.B.; Campbell, I.D.
J. Mol. Biol. 219, 717-725, 1991
A:Title: Three-dimensional structure of a complement control protein module in solut
A:Reference number: A49224; MUID:91278097
A:Contents: annotation: NMR structure determination, residues 927-985
R:Estaller, C.; Kolstien, V.; Schwaible, W.; Dierich, M.P.; Weiss, E.H.
J. Immunol. 146, 3190-3196, 1991
A:Title: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a n
A:Reference number: I56100; MUID:91201892
A:Accession: I72654
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1047-1231 <RES>
A:Cross-references: GB:M65294; NID:g183766; PIDN:AAA35948.1; PID:g183767
R:Carron, J.A.; Bates, R.C.; Smith, A.I.; Teto, T.; Arellano, A.; Gordon, D.L.; Bur
Biochim. Biophys. Acta 1289, 305-311, 1996
A:Title: Factor H co-purifies with thrombospondin isolated from platelet secretate.
A:Reference number: S66298; MUID:96205365

A:Accession: S66298
A:Status: preliminary
A:Molecule type: protein
A:Residues: 411-419;574-578;580-582 <AR>
C:Comment: Factor H has also been found bound to cell membranes in an unknown manner. HC
C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver.
C:Genetics: <HF1>
A:Gene: GDB:HF1; HF
A:Cross-references: GDB:120041; OMIM:134370
A:Map position: lq32-lq32
C:Genetics: <HF2>
A:Gene: GDB:HF2; HF
A:Cross-references: GDB:129095
A:Map position: lq32-lq32
A:Note: the correspondence between the two loci and the sequences indicated is unclear;
C:Function:
A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also increa
he alternative complement pathway
A:Pathway: complement alternate pathway
C:Superfamily: complement factor H; complement factor H repeat homology
C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma
F:1-18/Domain: signal sequence #status predicted <Sig>
F:19-129/Product: complement factor H #status experimental <MPT>
F:19-144/Product: complement factor H, short splice form #status experimental <MAT>
F:21-80/Domain: complement factor H repeat homology <FH01>
F:81-141/Domain: complement factor H repeat homology <FH02>
F:146-205/Domain: complement factor H repeat homology <FH03>
F:210-262/Domain: complement factor H repeat homology <FH04>
F:246-248/Region: cell attachment (R-G-D) motif
F:267-320/Domain: complement factor H repeat homology <FH05>
F:325-385/Domain: complement factor H repeat homology <FH06>
F:389-442/Domain: complement factor H repeat homology <FH07>
F:448-505/Domain: complement factor H repeat homology <FH08>
F:509-564/Domain: complement factor H repeat homology <FH09>
F:569-623/Domain: complement factor H repeat homology <FH10>
F:630-684/Domain: complement factor H repeat homology <FH11>
F:691-744/Domain: complement factor H repeat homology <FH12>
F:753-803/Domain: complement factor H repeat homology <FH13>
F:811-864/Domain: complement factor H repeat homology <FH14>
F:870-926/Domain: complement factor H repeat homology <FH15>
F:931-984/Domain: complement factor H repeat homology <FH16>
F:989-1043/Domain: complement factor H repeat homology <FH17>
F:1048-1102/Domain: complement factor H repeat homology <FH18>
F:1109-1163/Domain: complement factor H repeat homology <FH19>
F:1167-1228/Domain: complement factor H repeat homology <FH20>
F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357-
1-803,811-853,839-864,870-915,901-926,931-973,959-984,989-1032,1018-1043,1048-1091,1077-
F:529,802,822,882,911/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:718,1029,1095/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 1876; DB 1; Length 1231;
Best Local Similarity 100.0%; Pred. No. 3.3e-126;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 EDCNELPRRNTTEILTGSWSQDTPEGTQAIYKCRPGVRSIGNVIMVCRKGEWALNPLR 60
Db 19 EDCNELPRRNTTEILTGSWSQDTPEGTQAIYKCRPGVRSIGNVIMVCRKGEWALNPLR 78
Qy 61 KCQRKPCGHPGDTPTGTTFTLGGNVFEYGVKAVYTCNBYQLLGEINRYECDDTGTWNTDI 120
Db 79 KCQRKPCGHPGDTPTGTTFTLGGNVFEYGVKAVYTCNBYQLLGEINRYECDDTGTWNTDI 138
Qy 121 PICEVVKCLPVTAPENGKIVSSAMEPREYHFGQAVFVCSNGYKIEGDEMHCSDDGFW 180
Db 139 PICEVVKCLPVTAPENGKIVSSAMEPREYHFGQAVFVCSNGYKIEGDEMHCSDDGFW 198
Qy 181 SKEPKKVEISCKSPDVINGSPISQKLIYKENERFOYKCNNGYSEYERGDVAVCTESGWRP 240
Db 199 SKEPKKVEISCKSPDVINGSPISQKLIYKENERFOYKCNNGYSEYERGDVAVCTESGWRP 258
Qy 241 LPSCEKSCDNPYPNGDYSPLRIKHTGTDEITYQCRNGFYPATRGNTAKTSTGTWIPAP 300

Db 259 LPSCEKSCDNPYPNGDYSPLRIKHTGTDEITYQCRNGFYPATRGNTAKTSTGTWIPAP 318
Qy 301 RCTLKPCDDYPDIKKHGGLYHENMRPVPV 329
Db 319 RCTLKPCDDYPDIKKHGGLYHENMRPVPV 347
RESULT 3
NBMSH
complement factor H precursor - mouse
N:Alternate names: protein beta-1-H
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 22-Jun-1999
C:Accession: A26154; I49711; I49728
R:Kristensen, T.; Tack, B.F.
Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986
A:Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in lengt
A:Reference number: A26154; MUID:86233353
A:Accession: A26154
A:Molecule type: mRNA
A:Residues: 1-1234 <KRI>
R:Natsuume-Sakai, S.; Nonaka, M.; Nonaka, M.; Harada, Y.
J. Immunol. 144, 358-362, 1990
A:Title: Demonstration of an unusual allelic variation of mouse factor H by the compl
A:Reference number: I49711; MUID:90111033
A:Accession: I49711
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-18 <RES>
R:Cross-references: GB:M31979; NID:g193726; PIDN:AAA37762.1; PID:g193729
R:Munoz-Canoves, P.; Tack, B.F.; Vik, D.P.
Biochemistry 28, 9891-9897, 1989
A:Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma
A:Reference number: I49728; MUID:90148935
A:Accession: I49728
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-19 <RE2>
R:Cross-references: GB:J02891; NID:g193805; PIDN:AAA37795.1; PID:g553926
C:Comment: Two codominant alleles of factor H are present in mice.
C:Comment: Factor H functions as a cofactor in the inactivation of C3b by serine prot
(C5 convertase) in the alternative complement pathway.
C:Genetics:
A:Map position: 1
C:Superfamily: complement factor H; complement factor H repeat homology
C:Keywords: complement alternate pathway; duplication; glycoprotein; plasma
F:1-18/Domain: signal sequence #status predicted <Sig>
F:19-1234/Product: complement factor H #status predicted <MPT>
F:21-80/Domain: complement factor H repeat homology <FH01>
F:85-141/Domain: complement factor H repeat homology <FH02>
F:146-205/Domain: complement factor H repeat homology <FH03>
F:210-262/Domain: complement factor H repeat homology <FH04>
F:246-248/Region: cell attachment (R-G-D) motif
F:267-320/Domain: complement factor H repeat homology <FH05>
F:325-385/Domain: complement factor H repeat homology <FH06>
F:389-442/Domain: complement factor H repeat homology <FH07>
F:448-505/Domain: complement factor H repeat homology <FH08>
F:509-564/Domain: complement factor H repeat homology <FH09>
F:569-623/Domain: complement factor H repeat homology <FH10>
F:629-683/Domain: complement factor H repeat homology <FH11>
F:690-743/Domain: complement factor H repeat homology <FH12>
F:752-802/Domain: complement factor H repeat homology <FH13>
F:808-861/Domain: complement factor H repeat homology <FH14>
F:867-931/Domain: complement factor H repeat homology <FH15>
F:936-989/Domain: complement factor H repeat homology <FH16>
F:994-1048/Domain: complement factor H repeat homology <FH17>
F:1053-1107/Domain: complement factor H repeat homology <FH18>
F:1114-1168/Domain: complement factor H repeat homology <FH19>
F:1172-1233/Domain: complement factor H repeat homology <FH20>
F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,
0-802,808-850,836-861,867-920,906-931,936-978,964-989,994-1037,1023-1048,1053-1096,10

F:676,721,773,801,1030,1061,1225/Binding site: carbohydrate (Asn) (covalent) #status pre

Query Match 67.2%; Score 1261; DB 1; Length 1234;
Best Local Similarity 65.7%; Pred. No. 2.7e-82;
Matches 216; Conservative 37; Mismatches 76; Indels 0; Gaps 0;

QY 1 EDCNELPPRRNTTEILTGSNDSTYPEGTAIYKCRPGYRSLGNVIMVCRKGWVALNPLR 60
DB 19 EDCKGGPPPRENSEILLGSWSEQLYPEGTAIYKCRPGYRSLGNVIMVCRKGWVALNPLR 78

QY 61 KCQKRCCHGPGDTPGTFTLTGNGVFEYGVKAVYTCNEGQYLLGEINRECDTGTNDI 120
DB 79 ICRKKPCGHPGDPGSGFLAYGSEFQAKVYVTCDDGQYLLGEIDYRECGADGWINDI 138

QY 121 PICEVVKCLPVTAPENGKIVSSAMPEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFW 180
DB 139 PICEVVKCLPVTLENGRIVSAAETDQYIFGVVRFECNSGFKIEGHEIHCSENGLW 198

QY 181 SKEKPKCVESCKSPDVINGSPISOKIIYKENERFQYKCMGYEYSERGDVACTESGWRP 240
DB 199 SNEKPRCVETLTPRVENGDIIVKVPYKENERVHYCKHGVVPKERGDVACTESGWS 258

QY 241 LPSCEKSCDNPIYINGDYSPLRIKHTGDEITYQCRNGFYFATRGNTAKCTSTGWIPAP 300
DB 259 QPFCBEKSCSPYILNGIYTHRIHRSDDIETRYECNMGYFVPTGTVSKCTPTGWIPVP 318

QY 301 RCTLKPDCYDIKKGGLYHENRRPVPV 329
DB 319 RCTLKPCEFPQFKYGLYYEESLRNFPV 347

RESULT 4
S6551
factor H - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S6551
R:Soames, C.J.; Day, A.J.; Slim, R.B.
Biochem. J. 315, 523-531, 1996
A:Title: Prediction from sequence comparisons of residues of factor H involved in the
A:Reference number: S6551; MUID:96202005
A:Accession: S6551
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-669 <SO>
A:Cross-references: GB:X98697; NID:g1419423; PIDN:CAA67257.1; PID:g1419424
C:Superfamily: complement factor H; complement factor H repeat homology
F:55-114/Domain: complement factor H repeat homology <FHR1>
F:296-349/Domain: complement factor H repeat homology <FHR2>
F:355-412/Domain: complement factor H repeat homology <FHR3>
F:416-471/Domain: complement factor H repeat homology <FHR4>
F:476-530/Domain: complement factor H repeat homology <FHR5>
F:538-592/Domain: complement factor H repeat homology <FHR6>
F:599-651/Domain: complement factor H repeat homology <FHR7>

Query Match 48.1%; Score 903; DB 2; Length 669;
Best Local Similarity 60.9%; Pred. No. 5.2e-57;
Matches 154; Conservative 33; Mismatches 64; Indels 2; Gaps 1;

QY 76 GFTLTGNGVFEYGVKAVYTCNEGQYLLGEINRECDTGTNDIPICEVVKCLPVTAP 135
DB 3 GSPHLAENGQFYKAVYVTCDEGQYVGVVQVFEVVKCLPVTAP 62

QY 136 NGKIVSSAMEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFWSKKPKCVESCKSP 195
DB 63 NGKIFSDALEPDQYTYGVVQVFEVVKCLPVTAP 122

QY 196 DVINGSPISOKIIYKENERFQYKCMGYEYSERGDVACTESGWRP LPSCEKSCDNPIY 255
DB 123 VILNGQAVLPKATYKONERVQYRCAAGFEYGGQGTVCTKSGWTPAPTCTEITCDPPRP 182

QY 256 NGDYSPRLRIKHTRGDEITYQCRNGFYFATRGNTAKCTSTGWIPAPRCTILKPCDYPDIKHG 315
DB 183 NGYRPELSKYRGQDKITTECKKGFPEIRGTATCTRDGWWVPVPRCAWKPCSYPIVKG 242

QY 316 GLYHENMRPYP 328
DB 243 RLYYS--YRGYFP 253

RESULT 5
S6199
probable complement regulatory plasma protein SBI - barred sand bass
C:Species: Paralabrax nebulifer
C:Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 18-Jul-2001
C:Accession: S6199; S77894
R:Dahmen, A.; Kaidoh, T.; Zipfel, P.F.; Gigli, I.
Biochem. J. 301, 391-397, 1994
A:Title: Cloning and characterization of a cDNA representing a putative complement-r
A:Reference number: S6199; MUID:94318039
A:Accession: S6199
A:Molecule type: mRNA
A:Residues: 1-1053 <DAH1>
A:Cross-references: EMBL:L21703; NID:g639894; PIDN:AAA92556.1; PID:g639895
A:Experimental source: liver
A:Accession: S77894
A:Molecule type: protein
A:Residues: 526-532, 'X', 534-537: 809-817, 'X', 819-826 <DAH2>
C:Genetics:
A:Gene: SBI
C:Superfamily: complement factor H repeat homology
C:Keywords: glycoprotein
F:89-145/Domain: complement factor H repeat homology <FH01>
F:334-389/Domain: complement factor H repeat homology <FH02>
F:450-502/Domain: complement factor H repeat homology <FH03>
F:569-624/Domain: complement factor H repeat homology <FH04>
F:682-738/Domain: complement factor H repeat homology <FH05>
F:743-802/Domain: complement factor H repeat homology <FH06>
F:935-989/Domain: complement factor H repeat homology <FH07>
F:993-1052/Domain: complement factor H repeat homology <FH08>

Query Match 25.9%; Score 486; DB 2; Length 1053;
Best Local Similarity 29.7%; Pred. No. 5.1e-27;
Matches 114; Conservative 48; Mismatches 134; Indels 88; Gaps 13;

QY 21 DOTYPEGTAIYKCRPGYRSLGNVIMVCRKGWVALNPLRKCQKRCCHGPGDTPGTFTL 80
DB 47 EASYPGGRQVRVGCNNGYS--GFFKLVCYEGKWEIRG--AKQPRSCGHPGDAQFADFL 102

QY 81 TGNVFEYGVKAVYTCNEGQYLLGEINRECDTGTNDIPICEVVKCLPVTAPENGKIV 140
DB 103 AEGNDFVFGSKVYVTCQKGYQMVSRINRYRCVAEGWDGVVPCVCSQCCPLIHVDNNVQVI 162

QY 141 SSAMEPDREYHFGQAVRFVCSNGYKI--EGDEEMHCSDDGFWSKKPKCVESCKSPDVIN 199
DB 163 GG-----PEATFGNVVRFSCSRSEILDGSPELYCDERGDGSPVPCPKAKITCAIPPIEN 218

QY 200 GSPISOKIIYKENERFQYKCMGYEYSERGDVACTESG-----WRPLPSCEKSC----- 249
DB 219 GNVPAIRKYEKNDVLHYECDRAFKHDR--PSTCIKQIKAEWSPPTPLCESIKRCLTMD 277

QY 250 -----DNPYP-----NGOYS----- 260
DB 278 GTRYEPAYRNLFSPGETLKVICTARTSWISTPOETSVYVTTCDNGEMSIPTCQEVRCNSR 337

QY 261 -PLRI-----KRTGDEITYQCRNGFYFATRGNT--AKCTSTGWIPAPRCTILKPCDY 309
DB 338 REHVDSMDVRSWERYTDDNTRYCKRG--YKRTGGVTWATCGRNGWMPNPICEVTKCSK 396

QY 310 PDIKHG-----GLYHENMRPYP 326
DB 397 ENIQDAVIGTDKQIYNLNQKAIY 420

RESULT 6

S53711
C4BP alpha chain precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S53711
R:de Frutos, P.G.; Dahlbaeck, B. 1995
Biochim. Biophys. Acta 1261, 285-289, 1995
A:Title: cDNA structure of rabbit C4b-binding protein alpha-chain. Preserved sequence md
A:Reference number: S53711; MUID:95226458
A:Accession: S53711
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-597 <DEF>
A:Cross-references: EMBL:235490
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
F:50-107/Domain: complement factor H repeat homology <FH1>
F:112-169/Domain: complement factor H repeat homology <FH2>
F:174-234/Domain: complement factor H repeat homology <FH3>
F:239-294/Domain: complement factor H repeat homology <FH4>
F:299-360/Domain: complement factor H repeat homology <FH5>
F:364-422/Domain: complement factor H repeat homology <FH6>
F:426-480/Domain: complement factor H repeat homology <FH7>
F:484-538/Domain: complement factor H repeat homology <FH8>

Query Match 19.08; Score 356; DB 1; Length 597;
Best Local Similarity 30.5%; Pred. No. 5.3e-18;

Matches 106; Conservative 47; Mismatches 157; Indels 38; Gaps 20;

QY 2 DCNELPRRTEILTGSWSDQTYPEGTAIYKCRPGYRSLG-NVIMVCR-KGEWVALNPL 59
DB 49 DCGP-PHLLFASISLSENEYGTILKYTCRPGYRSLGILNPLTKCPRLW---SYD 104
QY 60 RKCKRPGCHGDPFPGFTLTGCVNFEYGVKAVYTCNEGYQLLGEINRYRCDTD----G 115
DB 105 TFCVKKRCRPGDLPGQVEVK--TDFSGQIEFSGEYILGST-SHCDLQEKVE 161
QY 116 WTNDIPICEVVKLPVTAPENGKIVSSAMEPDREYH-FGQAVRVCNSGYKIEDEEMHC 174
DB 162 WSDPLRCEIVKCEPPNPIINHHNGG---NEDIHYGSSVTVSCNPFSLLEASISFC 217
QY 175 S----DDGEVSKKPKVEISKSPDVINGSPISO-KIIYKENERFOYKCNMGVEYSERG 229
DB 218 TVRKNKTGVNWSPPVCKEILCSPNPNVPHGKIISGFGPIYKDSIMYTCIDGVL--RG 275
QY 230 DAV--C-TESGWRPLPS-CEEKSC-DNPIYIPNG-----DYSPLRIKHRTGDEITYQCRNG 279
DB 276 SSLHCELDKWNPPSPVCSNCLGLPNVPHASQOQYQWSTKEGVYVSGTELRYKCRPG 335
QY 280 FYP-ATRGNTAKC-TSGTWIPAPRCTLKPCDYPDIKIGGLYHNNRRP 325
DB 336 YRPVADEPIIVTQEDLRWSPFAGCEAICCPHPOLDNGAI-TEHSRNP 382

RESULT 7

NBUHCA
C4b-binding protein alpha chain precursor - human
N:Alternate names: C4BP; proline-rich protein
C:Species: Homo sapiens (man)
C:Date: 13-Aug-1986 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C:Accession: A33568; S02372; A90326; A24182; A93134; S29492; A31785; I52244; A03210
R:Matsuguchi, T.; Okamura, S.; Aso, T.; Sata, T.; Niho, Y.
Biochem. Biophys. Res. Commun. 165, 138-144, 1989
A:Title: Molecular cloning of the cDNA coding for proline-rich protein (PRP): identity c
A:Reference number: A33568; MUID:90073699
A:Accession: A33568
A:Molecule type: mRNA
A:Residues: 1-597 <MAL>
A:Cross-references: NID:g190501; PIDN:AAA36507.1; PID:g190502
A:Note: the authors translated the codon GGA for residue 492 as Glu
R:Lintin, S.J.; Lewin, A.R.; Reid, K.B.M.

FEBS Lett. 232, 328-332, 1988
A:Title: Derivation of the sequence of the signal peptide in human C4b-binding protei
A:Reference number: S02372; MUID:88242821
A:Accession: S02372
A:Molecule type: mRNA
A:Residues: 17-81 <LI2>
A:Cross-references: EMBL:X07853
A:Note: although the sequence determined extends to residue 9 above, these authors as
R:Chung, L.P.; Bentley, D.R.; Reid, K.B.M.
Biochem. J. 230, 133-141, 1985
A:Title: Molecular cloning and characterization of the cDNA coding for C4b-binding pr
A:Reference number: A90326; MUID:86025405
A:Accession: A90326
A:Molecule type: mRNA
A:Residues: 80-597 <CH2>
A:Cross-references: GB:X02865; NID:g29564; PIDN:CAA26617.1; PID:g29565
R:Lintin, S.J.; Reid, K.B.M.
FEBS Lett. 204, 77-81, 1986
A:Title: Studies on the structure of the human C4b-binding protein gene.
A:Reference number: A24182; MUID:86301119
A:Accession: A24182
A:Molecule type: DNA
A:Residues: 203-288 <LIN>
A:Cross-references: EMBL:X04284; EMBL:X04296
R:Rodriguez de Cordoba, S.; Sanchez-Corral, P.; Rey-Campos, J.
J. Exp. Med. 173, 1073-1082, 1991
A:Title: Structure of the gene coding for the alpha polypeptide chain of the human co
A:Reference number: A43023; MUID:91217619
A:Contents: annotation; exon-intron boundaries
R:Chung, L.P.; Gagnon, J.; Reid, K.B.M.
Mol. Immunol. 22, 427-435, 1985
A:Title: Amino acid sequence studies of human C4b-binding protein: N-terminal sequenc
ced by cyanogen bromide treatment.
A:Reference number: A93134; MUID:85296001
A:Accession: A93134
A:Molecule type: protein
A:Residues: 49-81 <CHI>
A:Note: this paper reports amino-terminal sequences of the intact protein and of a N
R:Hessing, M.; Kanter, D.; Takeya, H.; van't Veer, C.; Hackeng, T.M.; Iwanaga, S.; B
FEBS Lett. 317, 228-232, 1993
A:Title: The region Ser(333)-Arg(356) of the alpha-chain of human C4b-binding protein
A:Reference number: S29492; MUID:93146164
A:Accession: S29492
A:Status: preliminary
A:Molecule type: protein
A:Residues: 381-404 <HES>
R:Suzuki, K.; Nishioka, J.
J. Biol. Chem. 263, 17034-17039, 1988
A:Title: Binding site for vitamin K-dependent protein S on complement C4b-binding pro
A:Reference number: A31785; MUID:89034204
A:Accession: A31785
A:Molecule type: protein
A:Residues: 495-505, 'X', 507-510, 'X', 512-515 <SUZ>
A:Note: this peptide appears to bind protein S
R:Dahlback, B.; Smith, C.A.; Muller-Eberhard, H.J.
Proc. Natl. Acad. Sci. U.S.A. 80, 3461-3465, 1983
A:Title: Visualization of human C4b-binding protein and its complexes with vitamin K-
A:Reference number: A93950; MUID:83221615
A:Contents: annotation; electron microscopy; three-dimensional structure; ligand bind
R:Aso, T.; Okamura, S.; Matsuguchi, T.; Sakamoto, N.; Sata, T.; Niho, Y.
Biochem. Biophys. Res. Commun. 174, 222-227, 1991
A:Title: Genomic organization of the alpha chain of the human C4b-binding protein gen
A:Reference number: I52244; MUID:91113199
A:Accession: I52244
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-597 <ASO>
A:Cross-references: GB:M62486; NID:g190498; PIDN:AAA36506.1; PID:g190500
C:Comment: C4BP controls the classical pathway of complement activation. It binds as
the degradation of the C4bC2a complex (C3 convertase) by dissociating the complement
C:Comment: C4BP occurs in plasma in two forms, both of which bind complement fragment
isulfide bonds. The beta chain binds the vitamin K-dependent plasma protein S. Amino

actors V and VIII.
C:Comment: The molecule has a central body supporting seven tentacles (alpha chains), ea
C:Genetics:
A:Gene: GDB:C4BPA
A:Cross-references: GDB:120568; OMIM:120830
A:Map position: 1q32-1q32
A:Introns: 48/1; 110/1; 143/2; 172/1; 236/1; 297/1; 362/1; 425/1; 482/1; 540/3
A:Complex: octamer of seven alpha chains and one beta chain
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C:Keywords: acute phase; chylomicron; complement pathway; duplication; glycoprotein; pla
F:1-48/Domain: signal sequence #status predicted <SIG>
F:49-59/Product: C4b-binding protein alpha chain status predicted <MAT>
F:50-108/Domain: complement factor H repeat homology <FH1>
F:113-170/Domain: complement factor H repeat homology <FH2>
F:175-234/Domain: complement factor H repeat homology <FH3>
F:239-294/Domain: complement factor H repeat homology <FH4>
F:299-360/Domain: complement factor H repeat homology <FH5>
F:364-422/Domain: complement factor H repeat homology <FH6>
F:381-404/Region: complement C4b binding #status predicted
F:426-480/Domain: complement factor H repeat homology <FH7>
F:484-538/Domain: complement factor H repeat homology <FH8>
F:221,506,528/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 18.8%; Score 353; DB 1; Length 597;
Best Local Similarity 29.3%; Pred. No. 8.7e-18;
Matches 100; Conservative 50; Mismatches 141; Indels 50; Gaps 21;

Qy 20 SDQTYPEGTQAIYKCRPGY-BSLGNVIMVCRK-GEWVALNPLRCKQKPCGHPGDPFGT 77
Db 67 TETRFKGTTLKYTCCLPGVSHSTQTUTCNDSGGEW-YNTF-CIYARCRHPGE----- 118
Qy 78 FTLTGGNV-----FEYGVKAVYTCNEGYYQLLGEINREC-----GTNDIPICEVYK 128
Db 119 --LRNQVEIKTDLFSQIEFSGEFLIGTTSR-CEVQDRGVGWSHPLPQCEIVK 175
Qy 129 LPVTAPENKIVSSAMEPDREYHFGQAVRFVNCNRYKTEGDEEMHCSD-----GFWSKEK 184
Db 176 KPPDPDRNGRSGE-----ENFAYGFSVYSCDPRFSLGHASISCTVENETIGVWRSP 231
Qy 185 PKCIVEICKSPDVINGSPISO-KIYKENERFYKCNMGYYSERGDV--C-TESGWRP 240
Db 232 PTCEKITCRKPDVSHGENVSGFGPIYNYKDTIVFKCKQGFVL--RGSSVIHCADSKNPP 289
Qy 241 L-PSCEEKSCDN-PYIPNGDYS--PLRIK---HRTGDEITYOCNRNGFYPAT-RGNATK- 291
Db 290 SPPACEPNSCINLPDIPHASWETYPRTKEDYVYVGTVLRYCHPGYKPTTDEPTVICQ 349
Qy 292 TSTGWIPAPRCTLKPCDYPDKHGLYHENMRP-----YF 327
Db 350 KNLRTYQGCALCCPEKLNNGEITQHRKSRPANHCYVF 390

RESULT 8
A45900
Complement C3b receptor type 2 long form precursor - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 08-Oct-1999
C:Accession: A45900; I48306
R:Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Weis, J.H.
J. Immunol. 144, 3581-3591, 1990
A:Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2 gene
A:Reference number: A45900; MUID:90229754
A:Accession: A45900
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-676 <KUR>
A:Cross-references: GB:M36470
A:Experimental source: clone 31-1
R:Kim, Y.O.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.M.; Holers, V.
J. Exp. Med. 181, 151-159, 1995
A:Title: Mouse complement regulatory protein Cr1/p65 uses the specific mechanisms of bo
A:Reference number: I48306; MUID:95105691

A:Accession: I48306
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 21-367 <RES>
A:Cross-references: EMBL:U17128; NID:9595980; PIDN:AAA78271.1; PID:9595982
C:Genetics:
A:Gene: Cr2
A:Introns: 80/1; 113/2; 142/1; 275/1; 333/1
C:Superfamily: complement factor H repeat homology
F:22-78/Domain: complement factor H repeat homology <FH01>
F:83-140/Domain: complement factor H repeat homology <FH02>
F:145-211/Domain: complement factor H repeat homology <FH03>
F:217-272/Domain: complement factor H repeat homology <FH04>
F:276-331/Domain: complement factor H repeat homology <FH05>
F:336-394/Domain: complement factor H repeat homology <FH06>
F:399-458/Domain: complement factor H repeat homology <FH07>
F:467-523/Domain: complement factor H repeat homology <FH08>
F:531-587/Domain: complement factor H repeat homology <FH09>
F:592-648/Domain: complement factor H repeat homology <FH10>

Query Match 18.8%; Score 353; DB 2; Length 676;
Best Local Similarity 28.9%; Pred. No. 1e-17;
Matches 93; Conservative 61; Mismatches 136; Indels 32; Gaps 17;

Qy 20 SDQTYPEGTQAIYKCRPGYSLGNVIMVCRKGEWVALNPLRCKQKPCGHPGDPFGT 78
Db 37 SDKSEFALGTWEYKCRPGYKFSFIITCLETSKW--SDAQOFCRKKPCMNPOEPLGVS 94
Qy 79 TLTGQNVFEYGVKAVYTCNEGYYQLLGEINREC-----DTGWTNDIPICEVYKLPVTAP 134
Db 95 HINTG--IEFGSTIYSCNQYRLIGD--SSATCIYSDNTVMNDNMPLEIPCESPPAI 151
Qy 135 ENKTVSSAMEPDREYHFGQAVRFVNCNRYKTEGDEEMHCSD-----GFWSKEK 184
Db 152 SNGDYFSSRD---SFFYGVWVYTYCHTGKKNREKFLDLVGEKSIYCTSKDNQVGIWNSPP 208
Qy 185 PKCV-EISCKSPDVINGSPISO-KIYKENERFYKCNMGYYSERGDVAVC-TESGWR-P 240
Db 209 PQCIPIRVKCPMEIENGLVESGFKHSFFLNDTVIFKCKSGFTMKGSRIACQPNKWSPP 268
Qy 241 LPSCEEKSCDNPIYIPNGDYSPLRIKHRTGDEITYOCNRNGFYPATRNTAKCTSG-WI-P 298
Db 269 LPTCFMGLCPQNLHGDYNNKDEFFSVGQKVSYTCNFG-YTLIGTLNLECTSLGTWSNT 327
Qy 299 APRCTLKPCD-YPD-IKHGGLY 318
Db 328 VPTCEVKSCDAIPNHLHGRVF 349

RESULT 9
I73012
Complement C3b/C4b receptor, membrane-bound form precursor - human
N:Alternate names: complement C3b/C4b receptor; complement receptor type 1 (CRI); su
N:Contains: complement C3b/C4b receptor, secreted form
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 24-Nov-1999 #text_change 21-Jul-2000
C:Accession: I73012; I56203; A47602; S03291; S03843; A28507; A24748; C24748
R:Vik, D.P.; Wong, W.W.
J. Immunol. 151, 6214-6224, 1993
A:Title: Structure of the gene for the F allele of complement receptor type 1 and se
A:Reference number: I56203; MUID:94065175
A:Accession: I73012
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-683, 'X', 685-1133, 'X', 1135-1471, 'X', 1473-2489 <VIK1>
A:Cross-references: GB:LI7418; NID:9306678; PIDN:AAB60695.1; PID:g451303
A:Accession: I56203
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-683, 'X', 685-894, 'A', 896-1000, 1451-1471, 'X', 1473-2489 <VIK2>
A:Cross-references: GB:LI7418; NID:9306678; PIDN:AAB60694.1; PID:9306680
R:Wong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.,

J. Exp. Med. 169, 847-863, 1989
A:Title: Structure of the human CRI gene. Molecular basis of the structural and quantitative
A:Reference number: A47602; MUID:89176869
A:Accession: A47602
A:Molecule type: DNA
A:Residues: 1-41 <WON>
A:CROSS-references: GB:X14893
R:Hourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.
J. Exp. Med. 168, 1255-1270, 1988
A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b receptor
A:Reference number: S03291; MUID:89010527
A:Accession: S03291
A:Molecule type: mRNA
A:Residues: 26-584 <HOU>
A:CROSS-references: EMBL:X14362; NID:g30197; PIDN:CAA32541.1; PID:g736240
R:Experimental source: Clone CRI-4
R:Klickstein, L.B.; Bartow, T.J.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T.
J. Exp. Med. 168, 1699-1717, 1988
A:Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4b receptor
A:Reference number: S03843; MUID:89035992
A:Accession: S03843
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-894, 'A', 896-1000, 1451-2064, 'I', 2066-2276, 'P', 2278-2299, 'H', 2301-2325, 'T', 2327-2333, 'G', 2335-2353, 'C', 2355-2373, 'A', 2375-2393, 'G', 2395-2413, 'C', 2415-2433, 'A', 2435-2453, 'G', 2455-2473, 'C', 2475-2493, 'G', 2495-2513, 'C', 2515-2533, 'A', 2535-2553, 'G', 2555-2573, 'C', 2575-2593, 'G', 2595-2613, 'C', 2615-2633, 'A', 2635-2653, 'G', 2655-2673, 'C', 2675-2693, 'G', 2695-2713, 'C', 2715-2733, 'A', 2735-2753, 'G', 2755-2773, 'C', 2775-2793, 'G', 2795-2813, 'C', 2815-2833, 'A', 2835-2853, 'G', 2855-2873, 'C', 2875-2893, 'G', 2895-2913, 'C', 2915-2933, 'A', 2935-2953, 'G', 2955-2973, 'C', 2975-2993, 'G', 2995-3013, 'C', 3015-3033, 'A', 3035-3053, 'G', 3055-3073, 'C', 3075-3093, 'G', 3095-3113, 'C', 3115-3133, 'A', 3135-3153, 'G', 3155-3173, 'C', 3175-3193, 'G', 3195-3213, 'C', 3215-3233, 'A', 3235-3253, 'G', 3255-3273, 'C', 3275-3293, 'G', 3295-3313, 'C', 3315-3333, 'A', 3335-3353, 'G', 3355-3373, 'C', 3375-3393, 'G', 3395-3413, 'C', 3415-3433, 'A', 3435-3453, 'G', 3455-3473, 'C', 3475-3493, 'G', 3495-3513, 'C', 3515-3533, 'A', 3535-3553, 'G', 3555-3573, 'C', 3575-3593, 'G', 3595-3613, 'C', 3615-3633, 'A', 3635-3653, 'G', 3655-3673, 'C', 3675-3693, 'G', 3695-3713, 'C', 3715-3733, 'A', 3735-3753, 'G', 3755-3773, 'C', 3775-3793, 'G', 3795-3813, 'C', 3815-3833, 'A', 3835-3853, 'G', 3855-3873, 'C', 3875-3893, 'G', 3895-3913, 'C', 3915-3933, 'A', 3935-3953, 'G', 3955-3973, 'C', 3975-3993, 'G', 3995-4013, 'C', 4015-4033, 'A', 4035-4053, 'G', 4055-4073, 'C', 4075-4093, 'G', 4095-4113, 'C', 4115-4133, 'A', 4135-4153, 'G', 4155-4173, 'C', 4175-4193, 'G', 4195-4213, 'C', 4215-4233, 'A', 4235-4253, 'G', 4255-4273, 'C', 4275-4293, 'G', 4295-4313, 'C', 4315-4333, 'A', 4335-4353, 'G', 4355-4373, 'C', 4375-4393, 'G', 4395-4413, 'C', 4415-4433, 'A', 4435-4453, 'G', 4455-4473, 'C', 4475-4493, 'G', 4495-4513, 'C', 4515-4533, 'A', 4535-4553, 'G', 4555-4573, 'C', 4575-4593, 'G', 4595-4613, 'C', 4615-4633, 'A', 4635-4653, 'G', 4655-4673, 'C', 4675-4693, 'G', 4695-4713, 'C', 4715-4733, 'A', 4735-4753, 'G', 4755-4773, 'C', 4775-4793, 'G', 4795-4813, 'C', 4815-4833, 'A', 4835-4853, 'G', 4855-4873, 'C', 4875-4893, 'G', 4895-4913, 'C', 4915-4933, 'A', 4935-4953, 'G', 4955-4973, 'C', 4975-4993, 'G', 4995-5013, 'C', 5015-5033, 'A', 5035-5053, 'G', 5055-5073, 'C', 5075-5093, 'G', 5095-5113, 'C', 5115-5133, 'A', 5135-5153, 'G', 5155-5173, 'C', 5175-5193, 'G', 5195-5213, 'C', 5215-5233, 'A', 5235-5253, 'G', 5255-5273, 'C', 5275-5293, 'G', 5295-5313, 'C', 5315-5333, 'A', 5335-5353, 'G', 5355-5373, 'C', 5375-5393, 'G', 5395-5413, 'C', 5415-5433, 'A', 5435-5453, 'G', 5455-5473, 'C', 5475-5493, 'G', 5495-5513, 'C', 5515-5533, 'A', 5535-5553, 'G', 5555-5573, 'C', 5575-5593, 'G', 5595-5613, 'C', 5615-5633, 'A', 5635-5653, 'G', 5655-5673, 'C', 5675-5693, 'G', 5695-5713, 'C', 5715-5733, 'A', 5735-5753, 'G', 5755-5773, 'C', 5775-5793, 'G', 5795-5813, 'C', 5815-5833, 'A', 5835-5853, 'G', 5855-5873, 'C', 5875-5893, 'G', 5895-5913, 'C', 5915-5933, 'A', 5935-5953, 'G', 5955-5973, 'C', 5975-5993, 'G', 5995-6013, 'C', 6015-6033, 'A', 6035-6053, 'G', 6055-6073, 'C', 6075-6093, 'G', 6095-6113, 'C', 6115-6133, 'A', 6135-6153, 'G', 6155-6173, 'C', 6175-6193, 'G', 6195-6213, 'C', 6215-6233, 'A', 6235-6253, 'G', 6255-6273, 'C', 6275-6293, 'G', 6295-6313, 'C', 6315-6333, 'A', 6335-6353, 'G', 6355-6373, 'C', 6375-6393, 'G', 6395-6413, 'C', 6415-6433, 'A', 6435-6453, 'G', 6455-6473, 'C', 6475-6493, 'G', 6495-6513, 'C', 6515-6533, 'A', 6535-6553, 'G', 6555-6573, 'C', 6575-6593, 'G', 6595-6613, 'C', 6615-6633, 'A', 6635-6653, 'G', 6655-6673, 'C', 6675-6693, 'G', 6695-6713, 'C', 6715-6733, 'A', 6735-6753, 'G', 6755-6773, 'C', 6775-6793, 'G', 6795-6813, 'C', 6815-6833, 'A', 6835-6853, 'G', 6855-6873, 'C', 6875-6893, 'G', 6895-6913, 'C', 6915-6933, 'A', 6935-6953, 'G', 6955-6973, 'C', 6975-6993, 'G', 6995-7013, 'C', 7015-7033, 'A', 7035-7053, 'G', 7055-7073, 'C', 7075-7093, 'G', 7095-7113, 'C', 7115-7133, 'A', 7135-7153, 'G', 7155-7173, 'C', 7175-7193, 'G', 7195-7213, 'C', 7215-7233, 'A', 7235-7253, 'G', 7255-7273, 'C', 7275-7293, 'G', 7295-7313, 'C', 7315-7333, 'A', 7335-7353, 'G', 7355-7373, 'C', 7375-7393, 'G', 7395-7413, 'C', 7415-7433, 'A', 7435-7453, 'G', 7455-7473, 'C', 7475-7493, 'G', 7495-7513, 'C', 7515-7533, 'A', 7535-7553, 'G', 7555-7573, 'C', 7575-7593, 'G', 7595-7613, 'C', 7615-7633, 'A', 7635-7653, 'G', 7655-7673, 'C', 7675-7693, 'G', 7695-7713, 'C', 7715-7733, 'A', 7735-7753, 'G', 7755-7773, 'C', 7775-7793, 'G', 7795-7813, 'C', 7815-7833, 'A', 7835-7853, 'G', 7855-7873, 'C', 7875-7893, 'G', 7895-7913, 'C', 7915-7933, 'A', 7935-7953, 'G', 7955-7973, 'C', 7975-7993, 'G', 7995-8013, 'C', 8015-8033, 'A', 8035-8053, 'G', 8055-8073, 'C', 8075-8093, 'G', 8095-8113, 'C', 8115-8133, 'A', 8135-8153, 'G', 8155-8173, 'C', 8175-8193, 'G', 8195-8213, 'C', 8215-8233, 'A', 8235-8253, 'G', 8255-8273, 'C', 8275-8293, 'G', 8295-8313, 'C', 8315-8333, 'A', 8335-8353, 'G', 8355-8373, 'C', 8375-8393, 'G', 8395-8413, 'C', 8415-8433, 'A', 8435-8453, 'G', 8455-8473, 'C', 8475-8493, 'G', 8495-8513, 'C', 8515-8533, 'A', 8535-8553, 'G', 8555-8573, 'C', 8575-8593, 'G', 8595-8613, 'C', 8615-8633, 'A', 8635-8653, 'G', 8655-8673, 'C', 8675-8693, 'G', 8695-8713, 'C', 8715-8733, 'A', 8735-8753, 'G', 8755-8773, 'C', 8775-8793, 'G', 8795-8813, 'C', 8815-8833, 'A', 8835-8853, 'G', 8855-8873, 'C', 8875-8893, 'G', 8895-8913, 'C', 8915-8933, 'A', 8935-8953, 'G', 8955-8973, 'C', 8975-8993, 'G', 8995-9013, 'C', 9015-9033, 'A', 9035-9053, 'G', 9055-9073, 'C', 9075-9093, 'G', 9095-9113, 'C', 9115-9133, 'A', 9135-9153, 'G', 9155-9173, 'C', 9175-9193, 'G', 9195-9213, 'C', 9215-9233, 'A', 9235-9253, 'G', 9255-9273, 'C', 9275-9293, 'G', 9295-9313, 'C', 9315-9333, 'A', 9335-9353, 'G', 9355-9373, 'C', 9375-9393, 'G', 9395-9413, 'C', 9415-9433, 'A', 9435-9453, 'G', 9455-9473, 'C', 9475-9493, 'G', 9495-9513, 'C', 9515-9533, 'A', 9535-9553, 'G', 9555-9573, 'C', 9575-9593, 'G', 9595-9613, 'C', 9615-9633, 'A', 9635-9653, 'G', 9655-9673, 'C', 9675-9693, 'G', 9695-9713, 'C', 9715-9733, 'A', 9735-9753, 'G', 9755-9773, 'C', 9775-9793, 'G', 9795-9813, 'C', 9815-9833, 'A', 9835-9853, 'G', 9855-9873, 'C', 9875-9893, 'G', 9895-9913, 'C', 9915-9933, 'A', 9935-9953, 'G', 9955-9973, 'C', 9975-9993, 'G', 9995-10013, 'C', 10015-10033, 'A', 10035-10053, 'G', 10055-10073, 'C', 10075-10093, 'G', 10095-10113, 'C', 10115-10133, 'A', 10135-10153, 'G', 10155-10173, 'C', 10175-10193, 'G', 10195-10213, 'C', 10215-10233, 'A', 10235-10253, 'G', 10255-10273, 'C', 10275-10293, 'G', 10295-10313, 'C', 10315-10333, 'A', 10335-10353, 'G', 10355-10373, 'C', 10375-10393, 'G', 10395-10413, 'C', 10415-10433, 'A', 10435-10453, 'G', 10455-10473, 'C', 10475-10493, 'G', 10495-10513, 'C', 10515-10533, 'A', 10535-10553, 'G', 10555-10573, 'C', 10575-10593, 'G', 10595-10613, 'C', 10615-10633, 'A', 10635-10653, 'G', 10655-10673, 'C', 10675-10693, 'G', 10695-10713, 'C', 10715-10733, 'A', 10735-10753, 'G', 10755-10773, 'C', 10775-10793, 'G', 10795-10813, 'C', 10815-10833, 'A', 10835-10853, 'G', 10855-10873, 'C', 10875-10893, 'G', 10895-10913, 'C', 10915-10933, 'A', 10935-10953, 'G', 10955-10973, 'C', 10975-10993, 'G', 10995-11013, 'C', 11015-11033, 'A', 11035-11053, 'G', 11055-11073, 'C', 11075-11093, 'G', 11095-11113, 'C', 11115-11133, 'A', 11135-11153, 'G', 11155-11173, 'C', 11175-11193, 'G', 11195-11213, 'C', 11215-11233, 'A', 11235-11253, 'G', 11255-11273, 'C', 11275-11293, 'G', 11295-11313, 'C', 11315-11333, 'A', 11335-11353, 'G', 11355-11373, 'C', 11375-11393, 'G', 11395-11413, 'C', 11415-11433, 'A', 11435-11453, 'G', 11455-11473, 'C', 11475-11493, 'G', 11495-11513, 'C', 11515-11533, 'A', 11535-11553, 'G', 11555-11573, 'C', 11575-11593, 'G', 11595-11613, 'C', 11615-11633, 'A', 11635-11653, 'G', 11655-11673, 'C', 11675-11693, 'G', 11695-11713, 'C', 11715-11733, 'A', 11735-11753, 'G', 11755-11773, 'C', 11775-11793, 'G', 11795-11813, 'C', 11815-11833, 'A', 11835-11853, 'G', 11855-11873, 'C', 11875-11893, 'G', 11895-11913, 'C', 11915-11933, 'A', 11935-11953, 'G', 11955-11973, 'C', 11975-11993, 'G', 11995-12013, 'C', 12015-12033, 'A', 12035-12053, 'G', 12055-12073, 'C', 12075-12093, 'G', 12095-12113, 'C', 12115-12133, 'A', 12135-12153, 'G', 12155-12173, 'C', 12175-12193, 'G', 12195-12213, 'C', 12215-12233, 'A', 12235-12253, 'G', 12255-12273, 'C', 12275-12293, 'G', 12295-12313, 'C', 12315-12333, 'A', 12335-12353, 'G', 12355-12373, 'C', 12375-12393, 'G', 12395-12413, 'C', 12415-12433, 'A', 12435-12453, 'G', 12455-12473, 'C', 12475-12493, 'G', 12495-12513, 'C', 12515-12533, 'A', 12535-12553, 'G', 12555-12573, 'C', 12575-12593, 'G', 12595-12613, 'C', 12615-12633, 'A', 12635-12653, 'G', 12655-12673, 'C', 12675-12693, 'G', 12695-12713, 'C', 12715-12733, 'A', 12735-12753, 'G', 12755-12773, 'C', 12775-12793, 'G', 12795-12813, 'C', 12815-12833, 'A', 12835-12853, 'G', 12855-12873, 'C', 12875-12893, 'G', 12895-12913, 'C', 12915-12933, 'A', 12935-12953, 'G', 12955-12973, 'C', 12975-12993, 'G', 12995-13013, 'C', 13015-13033, 'A', 13035-13053, 'G', 13055-13073, 'C', 13075-13093, 'G', 13095-13113, 'C', 13115-13133, 'A', 13135-13153, 'G', 13155-13173, 'C', 13175-13193, 'G', 13195-13213, 'C', 13215-13233, 'A', 13235-13253, 'G', 13255-13273, 'C', 13275-13293, 'G', 13295-13313, 'C', 13315-13333, 'A', 13335-13353, 'G', 13355-13373, 'C', 13375-13393, 'G', 13395-13413, 'C', 13415-13433, 'A', 13435-13453, 'G', 13455-13473, 'C', 13475-13493, 'G', 13495-13513, 'C', 13515-13533, 'A', 13535-13553, 'G', 13555-13573, 'C', 13575-13593, 'G', 13595-13613, 'C', 13615-13633, 'A', 13635-13653, 'G', 13655-13673, 'C', 13675-13693, 'G', 13695-13713, 'C', 13715-13733, 'A', 13735-13753, 'G', 13755-13773, 'C', 13775-13793, 'G', 13795-13813, 'C', 13815-13833, 'A', 13835-13853, 'G', 13855-13873, 'C', 13875-13893, 'G', 13895-13913, 'C', 13915-13933, 'A', 13935-13953, 'G', 13955-13973, 'C', 13975-13993, 'G', 13995-14013, 'C', 14015-14033, 'A', 14035-14053, 'G', 14055-14073, 'C', 14075-14093, 'G', 14095-14113, 'C', 14115-14133, 'A', 14135-14153, 'G', 14155-14173, 'C', 14175-14193, 'G', 14195-14213, 'C', 14215-14233, 'A', 14235-14253, 'G', 14255-14273, 'C', 14275-14293, 'G', 14295-14313, 'C', 14315-14333, 'A', 14335-14353, 'G', 14355-14373, 'C', 14375-14393, 'G', 14395-14413, 'C', 14415-14433, 'A', 14435-14453, 'G', 14455-14473, 'C', 14475-14493, 'G', 14495-14513, 'C', 14515-14533, 'A', 14535-14553, 'G', 14555-14573, 'C', 14575-14593, 'G', 14595-14613, 'C', 14615-14633, 'A', 14635-14653, 'G', 14655-14673, 'C', 14675-14693, 'G', 14695-14713, 'C', 14715-14733, 'A', 14735-14753, 'G', 14755-14773, 'C', 14775-14793, 'G', 14795-14813, 'C', 14815-14833, 'A', 14835-14853, 'G', 14855-14873, 'C', 14875-14893, 'G', 14895-14913, 'C', 14915-14933, 'A', 14935-14953, 'G', 14955-14973, 'C', 14975-14993, 'G', 14995-15013, 'C', 15015-15033, 'A', 15035-15053, 'G', 15055-15073, 'C', 15075-15093, 'G', 15095-15113, 'C', 15115-15133, 'A', 15135-15153, 'G', 15155-15173, 'C', 15175-15193, 'G', 15195-15213, 'C', 15215-15233, 'A', 15235-15253, 'G', 15255-15273, 'C', 15275-15293, 'G', 15295-15313, 'C', 15315-15333, 'A', 15335-15353, 'G', 15355-15373, 'C', 15375-15393, 'G', 15395-15413, 'C', 15415-15433, 'A', 15435-15453, 'G', 15455-15473, 'C', 15475-15493, 'G', 15495-15513, 'C', 15515-15533, 'A', 15535-15553, 'G', 15555-15573, 'C', 15575-15593, 'G', 15595-15613, 'C', 15615-15633, 'A', 15635-15653, 'G', 15655-15673, 'C', 15675-15693, 'G', 15695-15713, 'C', 15715-15733, 'A', 15735-15753, 'G', 15755-15773, 'C', 15775-15793, 'G', 15795-15813, 'C', 15815-15833, 'A', 15835-15853, 'G', 15855-15873, 'C', 15875-15893, 'G', 15895-15913, 'C', 15915-15933, 'A', 15935-15953, 'G', 15955-15973, 'C', 15975-15993, 'G', 15995-16013, 'C', 16015-16033, 'A', 16035-16053, 'G', 16055-16073, 'C', 16075-16093, 'G', 16095-16113, 'C', 16115-16133, 'A', 16135-16153, 'G', 16155-16173, 'C', 16175-16193, 'G', 16195-16213, 'C', 16215-16233, 'A', 16235-16253, 'G', 16255-16273, 'C', 16275-16293, 'G', 16295-16313, 'C', 16315-16333, 'A', 16335-16353, 'G', 16355-16373, 'C', 16375-16393, 'G', 16395-16413, 'C', 16415-16433, 'A', 16435-16453, 'G', 16455-16473, 'C', 16475-16493, 'G', 16495-16513, 'C', 16515-16533, 'A', 16535-16553, 'G', 16555-16573, 'C', 16575-16593, 'G', 16595-16613, 'C', 16615-16633, 'A', 16635-16653, 'G', 16655-16673, 'C', 16675-16693, 'G', 16695-16713, 'C', 16715-16733, 'A', 16735-16753, 'G', 16755-16773, 'C', 16775-16793, 'G', 16795-16813, 'C', 16815-16833, 'A', 16835-16853, 'G', 16855-16873, 'C', 16875-16893, 'G', 16895-16913, 'C', 16915-16933, 'A', 16935-16953, 'G', 16955-16973, 'C', 16975-16993, 'G', 16995-17013, 'C', 17015-17033, 'A', 17035-17053, 'G', 17055-17073, 'C', 17075-17093, 'G', 17095-17113, 'C', 17115-17133, 'A', 17135-17153, 'G', 17155-17173, 'C', 17175-17193, 'G', 17195-17213, 'C', 17215-17233, 'A', 17235-17253, 'G', 17255-17273, 'C', 17275-17293, 'G', 17295-17313, 'C', 17315-17333, 'A', 17335-17353, 'G', 17355-17373, 'C', 17375-17393, 'G', 17395-17413, 'C', 17415-17433, 'A', 17435-17453, 'G', 17455-17473, 'C', 17475-17493, 'G', 17495-17513, 'C', 17515-17533, 'A', 17535-17553, 'G', 17555-17573, 'C', 17575-17593, 'G', 17595-17613, 'C', 17615-17633, 'A', 17635-17653, 'G', 17655-17673, 'C', 17675-17693, 'G', 17695-17713, 'C', 17715-17733, 'A', 17735-17753, 'G', 17755-17773, 'C', 17775-17793, 'G', 17795-17813, 'C', 17815-17833, 'A', 17835-17853, 'G',

00
-----RRCQKRRPGG 68
: : | |

Db	73	LQLKGGPPPEIPFAVHDGSSFGSEYDLDAEGLSISKCLLNKRKNVAQWFGPDURCKARACP	132
Qy	69	HPGDTPTGCTFLTGNNVFYGVKAVTCEGYOLLCEINRECDTDG-WTNDIPICEVVK	127
Db	133	DPGDIENG---LREGDTFEPHHVYSCNPGFLLVGST-ROCSSNGEWTPANCKKATE	188
Qy	128	CLPVTAPENKGIYSSAMEPDREYHFGQAVRFVCNCSYKIEGDEEHMCHSDDGFSWSEKPKC	187
Db	189	CSRSPSSPLHGKVVGSSILT-----YQSVVYSCDHCYRLVGQVQRICLAEGITWGNPEPC	242
Qy	188	VEISCK-SPDVG-----SPI-----	203
Db	243	EEIRCSVLPTLPNGYIEGSETSFCAVAVFRCLMTMHEGASKAKMEDQWSAPIRCLA	302
Qy	204	SQKIYENERFOYK-----CNMGYEYSEGDVCTESGWRPLPSCSEKS	248
Db	303	SCRVPHQNGKIKDKSEGQLIASKVIVECNKQHEANTDERLICSNSTWSHVPCVSL	362
Qy	249	CDN--PYIPNGDYSPLRIKHRTGDEITYOCRNGEFYPATRGNTAKTSTGWI--PAPRCT	303
Db	363	CHNWPVPVPHARI--LFSKSSHGSIACYECNNGYHPENRNQIIKCLYGTEWKDGPMMKCL	420
Qy	304	LKPCDYPDIKHGGL	317
Db	421	PSWCEHPDSKTYGTL	434

RESULT 13
T42921
complement control protein homolog ccph - ateline herpesvirus 3 (strain 73)
C:Species: ateline herpesvirus 3
A:Variety: strain 73
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 04-Mar-2000
C:Accession: T42921
R:Albrecht, J.C.; Fleckenstein, B.
A:Submitted to the EMBL Data Library, August 1998
A:Description: Primary structure of the herpesvirus ateles genome.
A:Reference number: Z22274
A:Accession: T42921
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-360 <ALB>
A:Cross-references: EMBL:AF083424; PID:AAC95530.1
A:Experimental source: strain 73

```
Query Match      17.9% ; Score 335; DB 2; Length 360;  
Best Local Similarity 33.4%; Pred. No. 9.8e-17;  
Matches 103; Conservative 32; Mismatches 119; Indels 54; Gaps 20;
```

QY	8 PRRNTEI-----LTGSWSODTYPEGTOAIKYCRPGYSRLGNVI--MVCRKGEWVALNP 59 : :
Db	24 PKRNRVSLRVVNITN--SSGYPNGTTLQVTCRGKGY--IGROIQTIVCVANGNTVPN-- 77 : :
QY	60 RKCKQRCGHPGDTPFGFTLGTGVNFVEYGKVAYCTNEG YQLLG EINYREC --DTDG-- 115 : :
Db	78 -ECCRCRSTPADLLNGWTYT-VNLX-YGSVITYTCTGYQLLSPT--SSCLLPDGRV 133 : :
QY	116 -WNIDIPICEVVVKCLPYTAPENGKI VVSAMEPDREYH-FGOAVFVCSNGYK ---IEGDEE 171 : :
Db	134 NWTRPPICEITCKPPTTIANGTHNI-----KEYTTYLDVAVYSCNDETKLTTGPSS 188 : :
QY	172 MHCSDDGFW-SKEKP KVEISCKSPDVINGSPI SQII KYENERFOY---KC NMGYEYSE 227 : :
Db	189 KQSETGWVPDETCKCEFFVKCIPOVANGHVVRKT--SNMVQOYVNIKRDKGFR LQG 246 : :
QY	228 RGA VCTESGWRP-LPSCEEKS CD NP VPINGDY SPLRIKHRTGT DE I TY QC RN GFYPATRG 286 : : :
Db	247 ETNMCKNGWFPA LP ICEK -----PAPPRGD-----MPHD SGED T ST ----PSGRN 290 : : :
QY	287 NTAKCTST 294 :

DB 291 CNOQNTTS 298

RESULT 14
JC2054

complement regulatory protein, 512 antigen precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 21-Jul-2000
C:Accession: JC2054; PC2027
R:Sakurada, C.; Seno, H.; Dohi, N.; Takizawa, H.; Nonaka, M.; Okada, H.
Biochem. Biophys. Res. Commun. 198, 819-826, 1994
A:Title: Molecular cloning of the rat complement regulatory protein, 512 antigen.
A:Reference number: JC2054; MUID:94161746
A:Accession: JC2054
A:Molecule type: mRNA
A:Residues: 1-497 <SAQ>
A:Cross-references: GB:D42114; NID:g577641; PIDN:g987688
A:Accession: PC2027
A:Molecule type: protein
A:Residues: 39-51 <SA2>
C:Comment: This protein plays a critical role in protection against complement mediated
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C:Keywords: glycoprotein; transmembrane protein
F:1-36/Domain: signal sequence #status predicted <SIG>
F:37-488/Product: complement regulatory protein, 512 antigen #status predicted <MAT>
F:38-94/Domain: complement factor H repeat homology <PH1>
F:99-156/Domain: complement factor H repeat homology <FH2>
F:161-227/Domain: complement factor H repeat homology <FH3>
F:233-388/Domain: complement factor H repeat homology <FH4>
F:234-352/Domain: complement factor H repeat homology <FH5>
F:357-413/Domain: complement factor H repeat homology <FH6>
F:421-444/Domain: transmembrane #status predicted <TMN>
F:247, 331, 346, 450, 482, 483/Binding site: Carbohydrate (Asn) (covalent) #status predicted

```

Query Match      17.2%   Score 323;   DB 2;   Length 497;
Best Local Similarity 28.3%;   Pred. No. 9.9e-16;
Matches 95;   Conservative 55;   Mismatches 144;   Indels 42;   Gaps 21;

QY  21  DOTYPECTQAIYKCRPGYRSLGNVIMYCRKGEWALNPLRKCKOKRPGCHGDPDTFGFTTL 80
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   55  ESTFPVSTYKUYECPGYIKRQFSITCEVNSVMT--SPQDVCIKKOCETPLDPONGIVHV 112
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  81  TGGNVFYGKAVYTCNEGYSQQLLEINRYBCDDTD--GWTNDIPICEVVK-KLPVTAPEN 136
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   113 --NTDIRFGSIITYTCNEGYSRLIGSSAMCIIISQSVANDAEAPICESICEIPPSIP-N 169
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  137 GKIVSSAMEPDRE-YHFGQAVRFVCSNG-----YKIEGDEEMHCS-----DDGFWSEKXP 185
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   170 GDFPS-----PNREDFHYGMVTVYQCNTDARGKKLFNLVGEPSIHCITSIDGOVCWWSGPPP 225
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  186 KCVETIS-CKSPDVLINGSPIQ-KIIYKENERFOYKCNMGYEYSERGDVCTE-SGHRP-L 241
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   226 QCIELNKCTPPHVENAVIVSKNLSLRLDMVFECODGFMKKGDSVYVYGRSLNRWEPQL 285
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  242 PSC-EERKSCDN--PYIPNGD-YSPLEIKHRTGEITYTCRNGFYPATRGNTAK-CTSTG- 295
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   286 PSCFKVKSCGAFGLPNGHVFPQNL--QLGAKVTFCVNTGY--QLKGNSSSHCVLDGV 341
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  296 --W-IPAPRCTLKPCDYPDIKHGGLYHENRRPYF 327
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   342 ESTWNSSVPVCEQVICKLPDMMSGFQKGLQMKDDY 377
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15
I46001
C4b-binding protein alpha chain - bovine
C:Species: Bos primigenius taurus (cattle)
C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: I46001; S43190
R:Hilliard, A.; Thern, A.; Dahlback, B.
J. Immunol. 153, 4190-4199, 1994
A:Title: Bovine C4b binding protein. Molecular cloning of the alpha- and beta-chains

A:Reference number: I46001; MUID:95015909
A:Accession: I46001
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-610 <HIL>
A:Cross-references: EMBL:231693; NID:9469117; PIDN:CAA83498.1; PID:9469118
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
F:50-107/Domain: complement factor H repeat homology <FH1>
F:112-169/Domain: complement factor H repeat homology <FH2>
F:174-234/Domain: complement factor H repeat homology <FH3>
F:239-294/Domain: complement factor H repeat homology <FH4>
F:299-362/Domain: complement factor H repeat homology <FH5>
F:366-425/Domain: complement factor H repeat homology <FH6>
F:429-483/Domain: complement factor H repeat homology <FHR>
F:487-541/Domain: complement factor H repeat homology <FH8>

Query Match 17.1%; Score 320.5; DB 1; Length 610;
Best Local Similarity 27.7%; Pred. No. 1.9e-15;
Matches 99; Conservative 54; Mismatches 142; Indels 63; Gaps 22;

QY 6 LPPRREILTGSWSDQTYPEGTQAIYKRCGYR--SLGNVIMVCRKGWVALNPLRKQ 63
 :| :
Db 52 IPPYLDFAFPINELNETFEFTGLTYTCRCGYRISRKNFLLICDGTDNW---KYKEFCV 108
 :| :
QY 64 KRPCGHPGDTFFGFTLTGGNV-----FEYGVKAVYTCNEGYQLGEIN-YRECDTG-- 115
 :| :
Db 109 KRCENPGE-----LLNGQVIVKTDYDFGSEIEFSCSEGVVLIGSANSYCOLQDKGV 161
 :| :
QY 116 WTNDIPICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVRFVCNSGYKIEGDEMHCS 175
 :| :
Db 162 WSDPLPQCIIAKCEPPPTISNGR--HNGGDED-FYTYGSSVYSCDRDFSMGRASISCR 218
 :| :
QY 176 DD---GFWSKEKPKCIVEISKSPDIVNGSPISQKI-----IYKENERFOYKCNMGVEY 225
 :| :
Db 219 VENKTIGVWSFSPSCCKKVICQPVVKG-----KITSGFGPIYTYQOSIYVACNKGRL 273
 :| :
QY 226 SERGDV--C-TEGWR-PLPSCSEKSC-DNPYPNGDYSPLRKHKRTGDE----- 271
 :| :
Db 274 --EGDSLIIHCEADNSWNPPTCELNGCLGLPHIPHALWE--RYDHTQTQTEQQVYDIGFV 329
 :| :
QY 272 ILYOCRNGFYPATRG-NTAKTST-GWIPAPRCTLKPCDYPDIKHGG--LYHENMRRP 325
 :| :
Db 330 LSYKCHFQYKPKETDPTTTCOSNLEWSPYIECKEVCCPEPNLNNYGSITLH---RRP 384
 :| :

Search completed: August 29, 2002, 15:02:17
Job time: 88 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:12:04 ; Search time 21.93 Seconds*
(without alignments)
580.882 Million cell updates/sec

Title: US-09-316-163-11
Perfect score: 1876
Sequence: 1 EDCNELPPRRNTEILTGSWS.....PDIKHGGLYHENRRPYPVP 329

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1876	100.0	1231	1	CFAH_HUMAN
2	1261	67.2	1234	1	CFAH_MOUSE
3	927.5	49.4	685	1	CFAH_BOVIN
4	353	18.8	597	1	C4BP_HUMAN
5	350.5	18.7	2039	1	CR1_HUMAN
6	320.5	17.1	610	1	C4BP_BOVIN
7	317	16.9	558	1	C4BP_RAT
8	311.5	16.6	489	1	C4BP_MOUSE
9	307	16.4	1033	1	CR2_HUMAN
10	303.5	16.2	1025	1	CR2_MOUSE
11	303	16.2	661	1	F13B_HUMAN
12	301.5	16.1	263	1	VCP_VACCV
13	294	15.7	668	1	F13B_MOUSE
14	286	15.2	830	1	LEM3_HUMAN
15	279.5	14.9	345	1	APOH_MOUSE
16	279	14.9	610	1	LEM2_HUMAN
17	276.5	14.7	612	1	LEM2_MOUSE
18	274.5	14.6	768	1	LEM3_MOUSE
19	274	14.6	377	1	MCP_HUMAN
20	274	14.6	768	1	LEM3_RAT
21	270.5	14.4	611	1	LEM2_CANFA
22	269.5	14.4	360	1	CCPH_HSVSA
23	267.5	14.3	345	1	APOH_HUMAN
24	266	14.2	484	1	LEM2_PIG
25	262.5	14.0	958	1	HIG_DROME
26	262	14.0	381	1	DAF_HUMAN
27	261.5	13.9	345	1	APOH_BOVIN
28	259	13.8	769	1	LEM3_SHEEP
29	252	13.4	340	1	DAF_PONPY
30	246.5	13.1	507	1	DAF_CAVPO
31	242.5	12.9	345	1	APOH_CANFA
32	241	12.8	549	1	LEM2_RAT
33	239.5	12.8	551	1	LEM2_RABIT

34 235 12.5 390 1 DAF1_MOUSE
35 233.5 12.4 485 1 LEM2_BOVIN
36 227.5 12.1 331 1 FHR4_HUMAN
37 227 12.1 330 1 FHR3_HUMAN
38 222.5 11.9 330 1 FHR1_HUMAN
39 222.5 11.9 407 1 DAF2_MOUSE
40 222 11.8 646 1 LEM3_BOVIN
41 217 11.6 297 1 APOH_RAT
42 210 11.2 372 1 LEM1_MOUSE
43 203.5 10.8 372 1 LEM1_HUMAN
44 203.5 10.8 372 1 LEM1_PANTR
45 202.5 10.8 372 1 LEM1_PONPY

Q61475 mus musculus
P98107 bos taurus
Q92496 homo sapien
Q02985 homo sapien
Q03591 homo sapien
Q61476 mus musculus
P42201 bos taurus
P25644 rattus norv
P18337 mus musculus
P14151 homo sapien
Q95237 pan troglod
Q95235 pongo pygma

ALIGNMENTS

RESULT 1
CFAH_HUMAN
ID CFAH_HUMAN STANDARD; PRT; 1231 AA.
AC P08603; Q14570; P78435; Q9NU86;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement factor H precursor (H factor I).
DE HFI OR HF OR CFH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT Y-402.
RC TISSUE=Liver;
RX MEDLINE=88134059; PubMed=2963625;
RA Ripoché J., Day A.J., Harris T.J.R., Sim R.B.;
RT "The complete amino acid sequence of human complement factor H.";
RL Biochem. J. 249:593-602(1988).
RN [2]
RP SEQUENCE OF 53-445 FROM N.A.
RX MEDLINE=87054207; PubMed=2946589;
RA Schulz T.F., Schwaible W., Stanley K.K., Weiss E., Dierich M.P.;
RT "Human complement factor H: isolation of cDNA clones and partial cDNA sequence of the 38-kDa tryptic fragment containing the binding site for C3b.";
RL Eur. J. Immunol. 16:1351-1355(1986).
RN [3]
RP SEQUENCE OF 226-449 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=86169701; PubMed=2937845;
RA Kristensen T., Wetsel R.A., Tack B.F.;
RT "Structural analysis of human complement protein H: homology with C4b binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";
RL J. Immunol. 136:3407-3411(1986).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX Bird C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1047-1231 FROM N.A.
RX MEDLINE=91201892; PubMed=1826708;
RA Estaller C., Koistinen V., Schwaible W., Dierich M.P., Weiss E.H.;
RT "Cloning of the 1.4-kb mRNA species of human complement factor H reveals a novel member of the short consensus repeat family related to the carboxy terminal of the classical 150-kDa molecule.";
RL J. Immunol. 146:3190-3196(1991).
RN [6]
RP SEQUENCE OF 19-35.
RX MEDLINE=83048213; PubMed=6215918;
RA Sim R.B., Discipio R.G.;
RT "Purification and structural studies on the complement-system control protein beta 1H (Factor H).";
RL Biochem. J. 205:285-293(1982).
RN [7]
RP SEQUENCE OF 1-19 FROM N.A.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
DR EMBL; M12660; AAA37759.1; -;
DR EMBL; J02891; AAA37795.1; -;
DR EMBL; M31979; AAA37762.1; -;
DR PIR; A26154; NBMNH.
DR HSP; P08603; IHFI.
DR GSD; MG1; 88385; Cfh.
DR InterPro; IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00084; sushi; 20.
DR SMART; SM00032; CCP; 20.
KW Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;
KW Signal.
FT SIGNAL 1 18
FT CHAIN 19 1234 BY SIMILARITY.
FT CHAIN COMPLEMENT FACTOR H.
FT DOMAIN 20 81 SUSHI 1.
FT DOMAIN 84 142 SUSHI 2.
FT DOMAIN 145 206 SUSHI 3.
FT DOMAIN 209 263 SUSHI 4.
FT DOMAIN 266 321 SUSHI 5.
FT DOMAIN 324 386 SUSHI 6.
FT DOMAIN 388 443 SUSHI 7.
FT DOMAIN 447 506 SUSHI 8.
FT DOMAIN 508 565 SUSHI 9.
FT DOMAIN 568 623 SUSHI 10.
FT DOMAIN 628 684 SUSHI 11.
FT DOMAIN 689 744 SUSHI 12.
FT DOMAIN 751 803 SUSHI 13.
FT DOMAIN 807 862 SUSHI 14.
FT DOMAIN 866 932 SUSHI 15.
FT DOMAIN 935 990 SUSHI 16.
FT DOMAIN 993 1049 SUSHI 17.
FT DOMAIN 1052 1108 SUSHI 18.
FT DOMAIN 1113 1169 SUSHI 19.
FT DOMAIN 1171 1234 SUSHI 20.
FT DISULFID 21 66 BY SIMILARITY.
FT DISULFID 52 80 BY SIMILARITY.
FT DISULFID 85 129 BY SIMILARITY.
FT DISULFID 114 141 BY SIMILARITY.
FT DISULFID 146 192 BY SIMILARITY.
FT DISULFID 178 205 BY SIMILARITY.
FT DISULFID 210 251 BY SIMILARITY.
FT DISULFID 237 262 BY SIMILARITY.
FT DISULFID 267 309 BY SIMILARITY.
FT DISULFID 294 320 BY SIMILARITY.
FT DISULFID 325 374 BY SIMILARITY.
FT DISULFID 357 385 BY SIMILARITY.
FT DISULFID 389 431 BY SIMILARITY.
FT DISULFID 416 442 BY SIMILARITY.
FT DISULFID 448 494 BY SIMILARITY.
FT DISULFID 477 505 BY SIMILARITY.
FT DISULFID 509 553 BY SIMILARITY.
FT DISULFID 536 564 BY SIMILARITY.
FT DISULFID 569 610 BY SIMILARITY.
FT DISULFID 597 622 BY SIMILARITY.
FT DISULFID 629 672 BY SIMILARITY.
FT DISULFID 658 683 BY SIMILARITY.
FT DISULFID 690 732 BY SIMILARITY.
FT DISULFID 718 743 BY SIMILARITY.
FT DISULFID 752 791 BY SIMILARITY.
FT DISULFID 780 802 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 861 BY SIMILARITY.
FT DISULFID 867 920 BY SIMILARITY.
FT DISULFID 906 931 BY SIMILARITY.
FT DISULFID 936 978 BY SIMILARITY.

```
FT DISULFID 964 989 BY SIMILARITY.
FT DISULFID 994 1037 BY SIMILARITY.
FT DISULFID 1023 1048 BY SIMILARITY.
FT DISULFID 1053 1096 BY SIMILARITY.
FT DISULFID 1082 1107 BY SIMILARITY.
FT DISULFID 1114 1157 BY SIMILARITY.
FT DISULFID 1143 1168 BY SIMILARITY.
FT DISULFID 1172 1223 BY SIMILARITY.
FT DISULFID 1206 1233 BY SIMILARITY.
FT CARBOHYD 676 676 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 773 773 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1061 1061 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1225 1225 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1234 AA; 139082 MW; C5AC02F341B957F7 CRC64;

Query Match 67.2%; Score 1261; DB 1; Length 1234;
Best Local Similarity 65.7%; Pred. No. 8.3e-92;
Matches 216; Conservative 37; Mismatches 76; Indels 0; Gaps 0;

Qy 1 EDCNELPPRRNTEILTGSWSDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPLR 60
Db 19 EDCRKGPPRENSILGSWSEQLYEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPLR 78

Qy 61 KCKRCPGHPGDTPTGTTLTGNNVFEYGVKAVYTCNMGYQLLGEINRECDTGTNDI 120
Db 79 ICRKPCGHPGDTPTGTSFLAVGSQFEGAKVYVYTCDDGYQLLGEIDYRECGADGWINDI 138

Qy 121 PICEVVKCLPVTAPANGKIVSAMPEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFW 180
Db 139 PLCEVVKCLPVTLENGRIVSGAAETDQYEGVQVVRFECSNGYKIEGHEKHCSENGLW 198

Qy 181 SKEPKPCVEISCKSPDVINGSPISOKIYKENERFQYKCNMGYSESGDAVCTESGWRP 240
Db 199 SNEKPRCEVILCTPRVNGDGINVKYKENERHYKCKHGYPKRGDAVCTESGWS 258

Qy 241 LPSCEKSCNDPNYPNGDYSPLRKIKHRTGDEITYQCRNGFYPATRNTAKTCTGTGWPAP 300
Db 259 QPCECEKRCSPYILNGIYVTHRIHSRDEIRYECNMGYFVPGTGSTVSKCTPTGTWIPVP 318

Qy 301 RCTLPCDYPDKHGLGHENMRFPVP 329
Db 319 RCTLPCDYPDKHGLGHENMRFPVP 347

RESULT 3
CPAH_BOVIN
ID CFAH_BOVIN STANDARD; PRT; 685 AA.
AC Q28085;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Complement factor H (H factor 1) (Fragments).
GN HFI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE OF 17-685 FROM N.A., AND SEQUENCE OF 1-16.
RC TISSUE=Liver;
RX MEDLINE=96202005; PubMed=8615824;
RA Soames C.J., Day A.J., Sim R.B.;
RT "Prediction from sequence comparisons of residues of factor H involved
in the interaction with complement component C3b.";
RL Biochem. J. 315:523-531(1996).
CC C3b by factor I and also increases the rate of dissociation of the
CC C3b complex (C3 convertase) and the (C3b)NBB complex (C5
CC
```

```
CC convertase) in the alternative complement pathway (By similarity).
CC -!- SIMILARITY: CONTAINS AT LEAST 13 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X98697; CAA67257.1; -.
CC HSSP; P10998; 1VVD.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00084; sushi; 11.
CC SMART; SM00032; CCP; 11.
CC Complement alternate pathway; Plasma; Repeat; Sushi.
FT NON_TER 1 17
FT DOMAIN <17 67 SUSHI 2.
FT DOMAIN 70 131 SUSHI 3.
FT DOMAIN 134 188 SUSHI 4.
FT DOMAIN 191 246 SUSHI 5.
FT DOMAIN 249 309 SUSHI 6.
FT DOMAIN 311 366 SUSHI 7.
FT DOMAIN 370 429 SUSHI 8.
FT DOMAIN 431 488 SUSHI 9.
FT DOMAIN 491 547 SUSHI 10.
FT DOMAIN 553 609 SUSHI 11.
FT DOMAIN 614 668 SUSHI 12.
FT DOMAIN 675 >685 SUSHI 13.
FT DISULFID 79 66 BY SIMILARITY.
FT DISULFID 71 117 BY SIMILARITY.
FT DISULFID 103 130 BY SIMILARITY.
FT DISULFID 135 176 BY SIMILARITY.
FT DISULFID 162 187 BY SIMILARITY.
FT DISULFID 192 234 BY SIMILARITY.
FT DISULFID 219 245 BY SIMILARITY.
FT DISULFID 250 297 BY SIMILARITY.
FT DISULFID 280 308 BY SIMILARITY.
FT DISULFID 312 354 BY SIMILARITY.
FT DISULFID 339 365 BY SIMILARITY.
FT DISULFID 371 417 BY SIMILARITY.
FT DISULFID 400 428 BY SIMILARITY.
FT DISULFID 432 476 BY SIMILARITY.
FT DISULFID 459 487 BY SIMILARITY.
FT DISULFID 492 534 BY SIMILARITY.
FT DISULFID 520 546 BY SIMILARITY.
FT DISULFID 554 597 BY SIMILARITY.
FT DISULFID 583 608 BY SIMILARITY.
FT DISULFID 615 656 BY SIMILARITY.
FT DISULFID 642 667 BY SIMILARITY.
FT NON_TER 685 685
SQ SEQUENCE 685 AA; 77536 MW; 69FC9DC8D530E872 CRC64;

Query Match 49.4%; Score 927.5; DB 1; Length 685;
Best Local Similarity 50.6%; Pred. No. 8.3e-66;
Matches 166; Conservative 35; Mismatches 68; Indels 59; Gaps 2;

Qy 1 EDCNELPPRRNTEILTGSWSDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPLR 60
Db 1 EDCKEPPPRKETEILSVS-----
Qy 61 KCKRCPGHPGDTPTGTTLTGNNVFEYGVKAVYTCNMGYQLLGEINRECDTGTNDI 120
Db 19 -----GSPHLAEGNQFEGAKVYVYTCDDGYQLLGEIDYRECGADGWINDI 63

Qy 121 PICEVVKCLPVTAPANGKIVSAMPEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFW 180
Db 64 PICEVVKCLPVTPENGKIFSDALEPDQYTYGVQVQECNSGYMLDGPQKHCAGGVW 123

Qy 181 SKEPKPCVEISCKSPDVINGSPISOKIYKENERFQYKCNMGYSESGDAVCTESGWRP 240
```

```

Db 124 SAETPKCVEIFKPPVILNGQAVLPKAYIKQNERVOYRCAGFEQGORQDTCTKSGWTP 183
QY 241 LPSCEEKSCDNPYIPNGDYSPLRIKHRTGDEITYOCNRYFPATRGNTAKCTSTGWIPAP 300
Db 184 APTCEITCDPPRIPNGVYRPELSKYRGDKITYECKGFFPEIRGTDTATCTRDGWVPV 243
QY 301 RCTLKPCDYPDIKHGLYHENRRYPFP 328
Db 244 RCWAKPCSYPIKHGRLYYS--YRGYFP 269

RESULT 4
C4BP_HUMAN STANDARD; PRT: 597 AA.
AC P04003;
DT 23-OCT-1986 (Rel. 02, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C4b-binding protein alpha chain precursor (C4bp) (Proline-rich
DE protein) (PRP).
GN C4BPA OR C4BP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=90073699; PubMed=2590215;
RA Matsuguchi T., Okamura S., Aso T., Sata T., Niho Y.;
RT "Molecular cloning of the cDNA coding for proline-rich protein (PRP):
RT identity of PRP as C4b-binding protein.";
RL Biochem. Biophys. Res. Commun. 165:138-144(1989).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91113199; PubMed=1989602;
RA Aso T., Okamura S., Matsuguchi T., Sakamoto N., Sata T., Niho Y.;
RT "Genomic organization of the alpha chain of the human C4b-binding
RT protein gene.";
RL Biochem. Biophys. Res. Commun. 174:222-227(1991).
[3]
RP SEQUENCE OF 9-81 FROM N.A.
RX MEDLINE=88242821; PubMed=3378624;
RA Lintin S.J., Levin A.R., Reid K.B.M.;
RT "Derivation of the sequence of the signal peptide in human
RT C4b-binding protein and interspecies cross-hybridisation of the C4bp
RT cDNA sequence.";
RL FEBS Lett. 232:328-332(1988).
[4]
RP SEQUENCE OF 203-288 FROM N.A.
RX MEDLINE=86301119; PubMed=3017751;
RA Lintin S.J., Reid K.B.M.;
RT "Studies on the structure of the human C4b-binding protein gene.";
RL FEBS Lett. 204:77-81(1986).
[5]
RP SEQUENCE OF 80-597 FROM N.A.
RX MEDLINE=86025405; PubMed=3840370;
RA Chung L.P., Bentley D.R., Reid K.B.M.;
RT "Molecular cloning and characterization of the cDNA coding for C4b-
RT binding protein, a regulatory protein of the classical pathway of the
RT human complement system.";
RL Biochem. J. 230:133-141(1985).
[6]
RP SEQUENCE OF 49-88.
RX MEDLINE=85296001; PubMed=4033666;
RA Chung L.P., Gagnon J., Reid K.B.M.;
RT "Amino acid sequence studies of human C4b-binding protein: N-terminal
RT sequence analysis and alignment of the fragments produced by limited
RT proteolysis with chymotrypsin and the peptides produced by cyanogen
RT bromide treatment.";
RL Mol. Immunol. 22:427-435(1985).
[7]

```

```

RP ELECTRON MICROSCOPY, 3D-STRUCTURE, AND LIGAND BINDING.
RX MEDLINE=83221615; PubMed=6222381;
RA Dahlback B., Smith C.A., Mueller-Eberhard H.J.;
RT "Visualization of human C4b-binding protein and its complexes with
RT vitamin K-dependent protein S and complement protein C4b.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:3461-3465(1983).
CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S
CC AND WITH SERUM AMYLOID P COMPONENT.
CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
CC OF 3 POSSIBLE SORTS: A 570 kDa COMPLEX OF 7 ALPHA CHAINS AND 1
CC BETA CHAIN, A 530 kDa HOMOPOLYMER OF ALPHA CHAINS OR A 500 kDa
CC COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF
CC THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE
CC BINDING SITE FOR C4B AT THE END.
CC -1- TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.
CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL; M31452; AAA36507.1; -
DR EMBL; M62486; AAA36506.1; -
DR EMBL; M62475; AAA36506.1; JOINED.
DR EMBL; M62476; AAA36506.1; JOINED.
DR EMBL; M62477; AAA36506.1; JOINED.
DR EMBL; M62478; AAA36506.1; JOINED.
DR EMBL; M62479; AAA36506.1; JOINED.
DR EMBL; M62480; AAA36506.1; JOINED.
DR EMBL; M62481; AAA36506.1; JOINED.
DR EMBL; M62482; AAA36506.1; JOINED.
DR EMBL; M62484; AAA36506.1; JOINED.
DR EMBL; M62485; AAA36506.1; JOINED.
DR EMBL; X07853; CAA30701.1; -
DR EMBL; X04284; CAB51244.1; -
DR EMBL; X04296; CAA27839.1; -
DR EMBL; X02865; CAA26617.1; -
DR PIR; A33588; NBHUC4.
DR HSSP; P10998; LVVD.
DR MIM; I20830; -
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 8.
DR SMART; SM00032; CCP; 8.
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal;
KW Polymorphism.
FT SIGNAL 1 48
FT CHAIN 49 597
FT DOMAIN 49 109
FT DOMAIN 112 171
FT DOMAIN 174 235
FT DOMAIN 238 295
FT DOMAIN 298 361
FT DOMAIN 364 423
FT DOMAIN 425 481
FT DOMAIN 483 539
FT DISULFID 50 96
FT DISULFID 81 108
FT DISULFID 113 154
FT DISULFID 140 170
FT DISULFID 175 217
FT DISULFID 203 234
FT DISULFID 239 281
C4B-BINDING PROTEIN ALPHA CHAIN.
SUSHI 1.
SUSHI 2.
SUSHI 3.
SUSHI 4.
SUSHI 5.
SUSHI 6.
SUSHI 7.
SUSHI 8.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.

```



```
FT DISULFID 267 294 BY SIMILARITY.
FT DISULFID 299 348 BY SIMILARITY.
FT DISULFID 332 360 BY SIMILARITY.
FT DISULFID 2364 2387 BY SIMILARITY.
FT DISULFID 2365 2409 BY SIMILARITY.
FT DISULFID 2399 422 BY SIMILARITY.
FT DISULFID 426 468 BY SIMILARITY.
FT DISULFID 454 480 BY SIMILARITY.
FT DISULFID 484 525 BY SIMILARITY.
FT DISULFID 511 538 BY SIMILARITY.
FT DISULFID 546 546 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT DISULFID 558 558 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .).
FT VARIANT 92 92 Q -> T.
FT VARIANT 357 357 /FTId-VAR_001977.
FT VARIANT 473 473 /FTId-VAR_001978.
FT VARIANT 473 473 W -> L (IN DBNP:1801341).
FT SEQUENCE 597 AA: 67033 MW: 67E03F2EA85A16DD CRC64;

Query Match 18.8%; Score 353; DB 1; Length 597;
Best Local Similarity 29.3%; Pred. No. 1.3e-20;
Matches 100; Conservative 50; Mismatches 141; Indels 50; Gaps 21;

QY 20 SDQTYEGTQALYKCRPGY-RSLGNVIMVCRK-GEWVALNLRKCKQKPCGHPGDPFGCT 77
DB 67 TETREKGTGLTKLPGVRSHTQTLCNSDGEWV-YNTF-CIYKCRHPGE----- 118
QY 78 FTLTGQNV-----FYGKAVYTCNEGYQLLGEINRECDTD----GWTNDIPICEVVKC 128
DB 119 --LRNGQVEIKTDLFGSGQIERSCSEGFELIGTTSR-CEVDRGVGWSHPLPQCEIVKC 175
QY 129 LPTVAPENKIVSSAMEPDREYHFCQAVRFVCSNCKYKIEGDEMCSD--GWSREK 184
DB 176 KPPDIRNGRHSGE-----ENFAYGFSVTYSCDPFSLGLGHASISCTVENETIGVWRPSP 231
QY 185 PKCIVEISCKSPDIVNGSPISQ-KIIVKENERFQYKCNMGYEYSERGDV--C-TESGWRP 240
DB 232 PTCEKITCKPDSHGEMVSGFGPIYNYKDTIVFKCKQGFVL--RGSSVIHCDADSKNWP 289
QY 241 L-PSCEEKSCDN-PYIPNGDYS--PLRTK---HRTGDEITYCQNGFYFAT-RGNTAKC- 291
DB 290 SPSPACEPNINLPDIPHASMETYERPTKEDYVYVGVTLVRCHEFCYKPTTDEPTTVICQ 349
QY 292 TSTGWIPAPRCTLKCDYDPIKHGGLYHENMRP-----YF 327
DB 350 KNLRTWYOGCEALCCPEPKLNGEITQHRKSRPANHCYVF 390

RESULT 5
CRL_HUMAN
ID CRL_HUMAN STANDARD; PRT; 2039 AA.
AC P17927;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement receptor type 1 precursor (C3b/C4b receptor) (CD35
DE antigen)
GN CRI OR C3BR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89035992; PubMed=2972794;
RA Klickstein L.B., Bartow T.J., Miletic V., Rabson L.D., Smith J.A.,
FA Fearon D.T.;
RT *Identification of distinct C3b and C4b recognition sites in the
```

```
RT human C3b/C4b receptor (CRI, CD35) by deletion mutagenesis.";
RL J. Exp. Med. 168:1699-1717(1988).
RN [2]
RP SEQUENCE OF 503-2039 FROM N.A.
RX MEDLINE=87168191; PubMed=2951479;
RA Klickstein L.B., Wong W.W., Smith J.A., Weis J.H., Wilson J.G.,
RA Fearon D.T.;
RT *Human C3b/C4b receptor (CRI). Demonstration of long homologous
RT repeating domains that are composed of the short consensus repeats
RT characteristics of C3/C4 binding proteins.";
RL J. Exp. Med. 165:1095-1112(1987).
RN [3]
RP SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.
RX MEDLINE=86067975; PubMed=2933745;
RA Wong W.W., Klickstein L.B., Smith J.A., Weis J.H., Fearon D.T.;
RT *Identification of a partial cDNA clone for the human receptor for
RT complement fragments C3b/C4b.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:7711-7715(1985).
CC -!- FUNCTION: CRI, PRESENT ON ERYTHROCYTES, LEUKOCYTES, GLOMERULAR
CC PODOCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS, MEDIATES THE
CC BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE
CC ACTIVATED COMPLEMENT.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- POLYMORPHISM: CRI IS RESPONSIBLE FOR THE KNOX BLOOD GROUP SYSTEM.
CC -!- MISCELLANEOUS: SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A
CC LONG HOMOLOGOUS REPEAT (LHR). THE N-TERMINAL TWO SRCS OF LHR-A
CC CONTAINED A SITE DETERMINING C4 SPECIFICITY. AND THE N-TERMINAL
CC TWO SRCS OF LHR-B AND -C EACH HAD A SITE DETERMINING C3
CC SPECIFICITY.
CC -!- MISCELLANEOUS: THIS IS THE SEQUENCE OF THE F ALLOTYPE OF CRI.
CC -!- SIMILARITY: CONTAINS 30 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M11569; AAA52297.1; -
CC EMBL; M11617; AAA52298.1; -
CC EMBL; M11618; AAA52299.1; -
CC EMBL; Y00816; CAA68755.1; -
CC EMBL; X05309; CAA28933.1; -
CC PIR; A28507; A28507.
CC PIR; A24748; A24748.
CC PIR; B24748; B24748.
CC PIR; C24748; C24748.
CC PIR; S03843; S03843.
CC HSP; P08603; 1HFI.
CC MIM; 120620; -
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00084; sushi; 30.
CC SMART; SM00032; CCP; 30.
KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW Receptor; Sushi; Blood group antigen.
FT SIGNAL 1 41
FT CHAIN 42 2039 COMPLEMENT RECEPTOR TYPE 1.
FT DOMAIN 42 1971 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1972 1996 POTENTIAL.
FT DOMAIN 1997 2039 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 42 42 PYRROLIDONE CARBOXYLIC ACID (POTENTIAL).
FT DOMAIN 42 100 SUSHI A1.
FT DOMAIN 103 162 SUSHI A2.
FT DOMAIN 165 233 SUSHI A3.
FT DOMAIN 237 294 SUSHI A4.
FT DOMAIN 296 354 SUSHI A5.
FT DOMAIN 357 417 SUSHI A6.
FT DOMAIN 420 488 SUSHI A7.
FT DOMAIN 492 550 SUSHI B1.
FT DOMAIN 553 612 SUSHI B2.
```

```
FT DOMAIN 615 683 SUSHI B3.
FT DOMAIN 687 744 SUSHI B4.
FT DOMAIN 746 804 SUSHI B5.
FT DOMAIN 807 867 SUSHI B6.
FT DOMAIN 870 938 SUSHI B7.
FT DOMAIN 942 1000 SUSHI C1.
FT DOMAIN 1003 1062 SUSHI C2.
FT DOMAIN 1065 1133 SUSHI C3.
FT DOMAIN 1137 1194 SUSHI C4.
FT DOMAIN 1196 1254 SUSHI C5.
FT DOMAIN 1257 1317 SUSHI C6.
FT DOMAIN 1320 1388 SUSHI C7.
FT DOMAIN 1395 1453 SUSHI D1.
FT DOMAIN 1456 1515 SUSHI D2.
FT DOMAIN 1518 1586 SUSHI D3.
FT DOMAIN 1590 1647 SUSHI D4.
FT DOMAIN 1649 1707 SUSHI D5.
FT DOMAIN 1710 1770 SUSHI D6.
FT DOMAIN 1773 1841 SUSHI D7.
FT DOMAIN 1847 1905 SUSHI E1.
FT DOMAIN 1908 1966 SUSHI E2.
FT DISULFID 43 86 BY SIMILARITY.
FT DISULFID 73 99 BY SIMILARITY.
FT DISULFID 104 145 BY SIMILARITY.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 195 232 BY SIMILARITY.
FT DISULFID 238 280 BY SIMILARITY.
FT DISULFID 266 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 358 400 BY SIMILARITY.
FT DISULFID 386 416 BY SIMILARITY.
FT DISULFID 421 470 BY SIMILARITY.
FT DISULFID 450 487 BY SIMILARITY.
FT DISULFID 493 536 BY SIMILARITY.
FT DISULFID 523 549 BY SIMILARITY.
FT DISULFID 554 595 BY SIMILARITY.
FT DISULFID 581 611 BY SIMILARITY.
FT DISULFID 616 665 BY SIMILARITY.
FT DISULFID 645 682 BY SIMILARITY.
FT DISULFID 688 730 BY SIMILARITY.
FT DISULFID 716 743 BY SIMILARITY.
FT DISULFID 747 790 BY SIMILARITY.
FT DISULFID 776 803 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 866 BY SIMILARITY.
FT DISULFID 900 937 BY SIMILARITY.
FT DISULFID 943 986 BY SIMILARITY.
FT DISULFID 973 999 BY SIMILARITY.
FT DISULFID 1004 1045 BY SIMILARITY.
FT DISULFID 1031 1061 BY SIMILARITY.
FT DISULFID 1066 1115 BY SIMILARITY.
FT DISULFID 1095 1132 BY SIMILARITY.
FT DISULFID 1138 1180 BY SIMILARITY.
FT DISULFID 1166 1193 BY SIMILARITY.
FT DISULFID 1197 1240 BY SIMILARITY.
FT DISULFID 1226 1253 BY SIMILARITY.
FT DISULFID 1258 1300 BY SIMILARITY.
FT DISULFID 1286 1316 BY SIMILARITY.
FT DISULFID 1321 1370 BY SIMILARITY.
FT DISULFID 1350 1387 BY SIMILARITY.
FT DISULFID 1396 1439 BY SIMILARITY.
FT DISULFID 1426 1452 BY SIMILARITY.
FT DISULFID 1457 1498 BY SIMILARITY.
FT DISULFID 1484 1514 BY SIMILARITY.
FT DISULFID 1519 1568 BY SIMILARITY.
FT DISULFID 1548 1585 BY SIMILARITY.
FT DISULFID 1591 1633 BY SIMILARITY.
FT DISULFID 1619 1646 BY SIMILARITY.
FT DISULFID 1650 1693 BY SIMILARITY.
FT DISULFID 1679 1706 BY SIMILARITY.

FT DISULFID 1711 1753 BY SIMILARITY.
FT DISULFID 1739 1769 BY SIMILARITY.
FT DISULFID 1774 1823 BY SIMILARITY.
FT DISULFID 1803 1840 BY SIMILARITY.
FT DISULFID 1848 1891 BY SIMILARITY.
FT DISULFID 1877 1904 BY SIMILARITY.
FT DISULFID 1909 1952 BY SIMILARITY.
FT DISULFID 1938 1965 BY SIMILARITY.
FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 509 509 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 860 860 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 959 959 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1028 1028 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1152 1152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1310 1310 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1481 1481 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1504 1504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1540 1540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1605 1605 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1763 1763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1908 1908 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2039 AA; 223589 MW; B2FD29B6AD3C5EB7 CRC64;

Query Match 18.7%; Score 350.5; DB 1; Length 2039;
Best Local Similarity 29.3%; Pred. No. 8.4e-20;
Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;

QY 3 CNE---LPPRRNTEILTGSWSDQTYPEGTQAIYKCRPGYKSLGNVIMVCRKGWVALNPL 59
DB 43 CNAPEWLFPARPNLT----DEFEFGTYLNYECPGYSGRPFISICLNSVMTGARD- 97
QY 60 RKCOKRPGCHPGDTPFGTFTLTGNGVFEYGVKAVYTCNEGYQLLGEINYNREC---DTDG 115
DB 98 -RCRRKSCRNPDPVNGVHVING--IQGSGQIKYSGCTKGYRLIGS--SSATCIISGDTVI 153
QY 116 WTNDIPICEVVKC-LPVTAPENGIYSSAMEPDREYHFGQAVRFVCNSG-----YKIEG 168
DB 154 WDNETPICDRIPCGLPPTI-TNGDFISTNRE---NEHYGSVVTVYRCNPGSGGRKRVFLVG 209
QY 169 DEEMHC--SDD--GFWSKEKPKC-VEISCKSPDVINGSPIS-OKIYKENEREQYKCNMG 222
DB 210 EPSIYCTSDNDQVGVIWSPAPQCIIIPNKPCTPPNVENGILVSDNRSLFSLNEVYEFRCQP 269
QY 223 YEVSERGDVCTE-SGWRP-LPSCEEKSCDNPVYPN-----GDYSPLRKHRTG 269
DB 270 FVMKGPVRVKCALNKWPELPSC-SRVQPP--PDVLAERTQDKDNFSP-----G 319
QY 270 DEITYQCRNGFYPATRGNTAKTSTG-WIP-APRCTLKPCD 308
DB 320 QEVFYSCEPG-YDLRGAASMRCTPQGDWSPAAPTCEVKSCD 359

RESULT 6
C4BP-BOVIN
ID C4BP-BOVIN STANDARD; PRT; 610 AA.
AC Q28065;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C4b-binding protein alpha chain precursor (C4bp).
GN C4BPA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
```

Matches	99;	Conservative	54;	Mismatches	142;	Indels	63;	Gaps
QY	6	LPPRNRTEILTGSWSQDTYEGTOAIKYCRPGYR--SLGNVIMVKRGEMVALPLRKQC	63					
DB	52	IPPYLDFAFFINELNETRFETGTTLRVTCRPGYRISSRRKNFLDGDGMN---KVKEFCV	108					
QY	64	KRPCGHGRDPFGFTTLTGNNV-----FEYGKAVYCNCBEGYQLLGEIN-YRECDTDG--	115					
DB	109	KKRCENPGE-----LLNGQIVIKTYDSFSGSETFSCSEGVLIGSANSYCOLQDKGW	161					
QY	116	WTNDPIPCVVKLPTAPENGKIIVSAMPEPDHYHFGQAIFVCNSGYKTGEDDEMHCS	175					
DB	162	WSDLPLOCIATAKCEPPTISNGR--HNGGDED-FYTGGSVITYSCDRDFSMGLKRASISR	218					
QY	176	DD-----GFWSKEKPCKVEISCSPDVINGSPISOKI-----IYKENERFOYKCNMGYEY	225					
DB	219	VENKITIGWMSPPSPSKKKVICQPWWKD-----KITSGFGPIVTYQQSIIVACNKGFRL	273					
QY	226	SERGDAV-C-TESGWR-PLPSCEEKSC-DNPYIPNGDYSPLRIRKHRTGDE-----	271					
DB	274	--EGDSLIIHCADSNWNPPPTCELNGCLGLPHIPHALWE--RYDHQTGTQEQQYYDIGFV	329					
QY	272	ITCYCRNGFYPATRG-NIACTKTST-GWIAPRCTLKPCDYPDIDKHGG--LYHENMRRP	325					
DB	330	LSYKCHGYGPETDGTPTTVTCQSLEWSPYIECKEVCCPEPNLNYSITLH----RRP	384					
<p>RESULT 7 C4BP_RAT STANDARD; PRT; 558 AA.</p>								
ID	C4BP_RAT	Q63514;						
AC	Q63514;							
DT	01-NOV-1997 (Rel. 35, Created)							
DT	01-NOV-1997 (Rel. 35, Last sequence update)							
DT	10-OCT-2001 (Rel. 40, Last annotation update)							
DE	C4BP-binding protein alpha chain precursor (C4bp).							
DN	GABPA.							
GC	Rattus norvegicus (Rat).							
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.							
NCBI_TaxID=10116;	[1]							
SEQUENCE FROM N.A.								
STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;								
MEDLINE=97166082; PubMed=9013975;								
Hillarp A., Wiklund H., Thern A., Dahlback B.:								
"Molecular cloning of rat C4b binding protein alpha- and beta-chains: structural and functional relationships among human, bovine, rabbit, mouse, and rat proteins.";								
J. Immunol. 158:1315-1323(1997).								
-! FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT ACTIVATION. IT BINDS AS A COFACTORS TO C3B/C4B INACTIVATOR (C3BINA), WHICH THEN HYDROLYSES THE COMPLEMENT FRAGMENT C4B. IT ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3 CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S AND WITH SERUM AMYLLOYD P COMPONENT.								
-! SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.								
-! SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.								
-! SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.								
This SWISS-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).								
EMBL; 250051; CAA90391.1; .								
HSSP; F10996; 1LVVD.								
InterPro; IPR000436; Sushi_SCR_CCP.								
Pfam; PF00084; sushi; 8.								
SMART; SMO0032; SCR_0.								

KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT SIGNAL 1 13
FT CHAIN 14 558
FT DOMAIN 14 73
FT SUSHI 1. 14 73
FT DOMAIN 76 135
FT SUSHI 2. 76 135
FT DOMAIN 138 200
FT SUSHI 3. 138 200
FT DOMAIN 203 259
FT SUSHI 4. 203 259
FT DOMAIN 262 325
FT SUSHI 5. 262 325
FT DOMAIN 328 387
FT SUSHI 6. 328 387
FT DOMAIN 389 444
FT SUSHI 7. 389 444
FT DOMAIN 446 502
FT SUSHI 8. 446 502
FT DISULFID 15 60
FT BY SIMILARITY. 15 60
FT DISULFID 45 72
FT BY SIMILARITY. 45 72
FT DISULFID 77 118
FT BY SIMILARITY. 77 118
FT DISULFID 104 134
FT BY SIMILARITY. 104 134
FT DISULFID 139 182
FT BY SIMILARITY. 139 182
FT DISULFID 168 199
FT BY SIMILARITY. 168 199
FT DISULFID 204 246
FT BY SIMILARITY. 204 246
FT DISULFID 232 258
FT BY SIMILARITY. 232 258
FT DISULFID 263 312
FT BY SIMILARITY. 263 312
FT DISULFID 296 324
FT BY SIMILARITY. 296 324
FT DISULFID 7328 7351
FT BY SIMILARITY. 7328 7351
FT DISULFID 7329 7373
FT BY SIMILARITY. 7329 7373
FT DISULFID 7363 386
FT BY SIMILARITY. 7363 386
FT DISULFID 390 431
FT BY SIMILARITY. 390 431
FT DISULFID 417 443
FT BY SIMILARITY. 417 443
FT DISULFID 447 488
FT BY SIMILARITY. 447 488
FT DISULFID 474 501
FT BY SIMILARITY. 474 501
FT DISULFID 509 509
FT BY SIMILARITY. 509 509
FT DISULFID 521 521
FT BY SIMILARITY. 521 521
FT CARBOHYD 31 31
FT N-LINKED (GLCNAC. . .) (POTENTIAL). 31 31
FT CARBOHYD 177 177
FT N-LINKED (GLCNAC. . .) (POTENTIAL). 177 177
FT CARBOHYD 186 186
FT N-LINKED (GLCNAC. . .) (POTENTIAL). 186 186
FT CARBOHYD 469 469
FT N-LINKED (GLCNAC. . .) (POTENTIAL). 469 469
FT CARBOHYD 491 491
FT N-LINKED (GLCNAC. . .) (POTENTIAL). 491 491
SQ SEQUENCE 558 AA; 62266 MW; 592F0C667EDIE5FF CRC64;

Query Match 16.9%; Score 317; DB 1; Length 558;
Best Local Similarity 27.0%; Pred. No. 8.6e-18;
Matches 110; Conservative 46; Mismatches 128; Indels 124; Gaps 24;

Qy 7 PPRNTEILTGSWSQDT-YPEQTOAIYKCRPGY-RSLGNVIMVCRK-GEW---VALNPLR 60
Db 17 PPPDLPALPASEMNTDFESHITLRYNCRPGYSRASSQSLYKPLKQWQINIA----- 71

Qy 61 KCKRPCGHPGDPRTGTTLTGGNV-----FEYGVKAVYTCNEGQLLG-EINYRECDTD 114
Db 72 -CVKSCRNPGD-----LQNGKVEVKTDFLFGSQIEFSCSEGVILIGSSTSYCEIQCK 123

Qy 115 G--WTNDIPICEVVKCLPVTAPENCK-----IVSSAMEPD----- 147
Db 124 GVSWSDPLECVIAKCGMPDINSCKHNGREEEFTYRSVYTKCDPDTLLGNASITCT 183

Qy 148 -----REYHFGQAVRFVNCNGYKLEGDEE 171
Db 184 VVNTGVWSPSPPCERIIICPWPVKLVHCTINSFKHYKYKDSVRFQCGFVLURSGV 243

Qy 172 MHCSDGFWSEKPKVCBEISCKS-PDVINGSPIS-----QKIYKENERFOYKCMNGY 225
Db 244 IHCADGWS-PVPVCELNSCTDIPDIPNAALITSPRKEDVYPVGVTLVYICRPGYEP 302

Qy 226 SERG-DAVCTES-GWRPLPSCSEKSCNDVPYINGDYSPLR-IKHRT-----GD 270
Db 303 ATROPMTVICQKDLWSMLRCKEICCVP-----DPKSVRVIQHEKAHPNDQCTYFRGD 357

Qy 271 EITYQCRNGFYPATRGNTAKTSTG-WIP-APRCTLKPDPYD-IKHG 315
Db 358 EVSYTCQNDIN-----LTATCKSDGTHWPTPSCH-QSCDPPPAIAHG 399

RESULT 8
C4BP_MOUSE

ID C4BP_MOUSE STANDARD; PRT: 469 AA.
AC P08607;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C4B-binding protein precursor (C4bp).
GN C4BPA OR C4BP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88024997; PubMed=3663616;
RA Kristensen T., Ogata R.T., Chung L.P., Reid K.B.M., Tack B.F.;
RT *cDNA structure of murine C4b-binding protein, a regulatory component
of the serum complement system.*;
RL Biochemistry 26:4668-4674(1987).
CC -|- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
(C3BIINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
-|- SUBUNIT: HOMOHEPTAMER; NOT COVALENTLY LINKED. MOUSE LACKS THE
BETA CHAIN OF C4BP.
CC -|- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -|- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS
CC -|- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-44 IS THE INITIATOR.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
EMBL; M17122; AAA37312.1; ALT_INIT.
DR PIR; A27117; NBMSC4.
DR HSSP; P10998; 1VVD.
DR MGD; MGI:88229; C4bp.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 6.
DR SMART; SM00032; CCP; 6.
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT SIGNAL 1 56
FT CHAIN 57 469
FT C4B-BINDING PROTEIN.
FT DOMAIN 57 116
FT SUSHI 1.
FT DOMAIN 119 177
FT SUSHI 2.
FT DOMAIN 180 241
FT SUSHI 3.
FT DOMAIN 244 300
FT SUSHI 4.
FT DOMAIN 302 356
FT SUSHI 5.
FT DOMAIN 358 414
FT SUSHI 6.
FT DISULFID 58 103
FT BY SIMILARITY.
FT DISULFID 88 115
FT BY SIMILARITY.
FT DISULFID 120 160
FT BY SIMILARITY.
FT DISULFID 146 176
FT BY SIMILARITY.
FT DISULFID 181 223
FT BY SIMILARITY.
FT DISULFID 209 240
FT BY SIMILARITY.
FT DISULFID 245 287
FT BY SIMILARITY.
FT DISULFID 273 299
FT BY SIMILARITY.
FT DISULFID 303 343
FT BY SIMILARITY.
FT DISULFID 329 355
FT BY SIMILARITY.
FT DISULFID 359 400
FT BY SIMILARITY.
FT DISULFID 386 413
FT BY SIMILARITY.
FT CARBOHYD 74 74
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 381 381
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 428 428
FT N-LINKED (GLCNAC. . .) (POTENTIAL).

[illegible]

```
FT DISULFID 118 146 BY SIMILARITY.
FT DISULFID 134 197 BY SIMILARITY.
FT DISULFID 183 210 BY SIMILARITY.
FT DISULFID 215 256 BY SIMILARITY.
FT DISULFID 242 271 BY SIMILARITY.
FT DISULFID 276 325 BY SIMILARITY.
FT DISULFID 305 342 BY SIMILARITY.
FT DISULFID 351 393 BY SIMILARITY.
FT DISULFID 379 406 BY SIMILARITY.
FT DISULFID 410 453 BY SIMILARITY.
FT DISULFID 439 466 BY SIMILARITY.
FT DISULFID 471 509 BY SIMILARITY.
FT DISULFID 495 522 BY SIMILARITY.
FT DISULFID 527 576 BY SIMILARITY.
FT DISULFID 536 593 BY SIMILARITY.
FT DISULFID 602 644 BY SIMILARITY.
FT DISULFID 630 657 BY SIMILARITY.
FT DISULFID 662 699 BY SIMILARITY.
FT DISULFID 685 714 BY SIMILARITY.
FT DISULFID 719 762 BY SIMILARITY.
FT DISULFID 748 779 BY SIMILARITY.
FT DISULFID 788 830 BY SIMILARITY.
FT DISULFID 816 843 BY SIMILARITY.
FT DISULFID 851 894 BY SIMILARITY.
FT DISULFID 880 907 BY SIMILARITY.
FT DISULFID 912 955 BY SIMILARITY.
FT DISULFID 941 968 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 127 127 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 492 492 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 682 682 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 800 800 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 823 823 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 861 861 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 911 911 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 949 949 MISSING (IN SHORT ISOFORM).
FT VARSPLIC 525 556 ITCPPPPVYNGAHTGSSLEDPFGTGYTYTC -> NHLPT
FT TPCLYWGTHREFLRFSIWNHGLHM (IN SHORT
FT ISOFORM).
FT CONFLICT 667 667 Q -> D (IN REF. 2).
FT CONFLICT 902 902 Q -> G (IN REF. 2).
FT CONFLICT 906 906 H -> L (IN REF. 2).
FT SEQUENCE 1033 AA; 112973 MW; 1749DBA407847ADA CRC64;

Query Match 16.4%; Score 307; DB 1; Length 1033;
Best Local Similarity 23.7%; Pred. No. 1e-16;
Matches 97; Conservative 58; Mismatches 139; Indels 116; Gaps 22:

QY 1 EDCNELPPRRNTEILTGSWSDQ---TYPEGTQAIYKCRPGYRSLGNVIMVC-RKGEWVAL 56
Db 408 REC-QAPP---NILNGQKEDRHMYRDPGTSIKYSCNPGYVLVEESIQCTSEGWVTP- 461
QY 57 NPLRCQRPCGHPG---DTPFGFTTLTGNVFEYGVKAVYTCNEGVLQGEINRYRED 112
Db 462 -VPQCKVAACEATGRQLLTQPHQFVRPDVN-----SSCGEYKLSGSV-YQECQ 510
QY 113 -TDGNTNDIPICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVRFVCSNG-----YK 165
Db 511 GTIPWFMEIRLCKEITCPPPPVYNGAHTGSSLE---DFPYGTGYTYTCNPGPERGVFEFS 567
QY 166 IEGDEMHCSDD---GFWSEKPKPC-----VEISCKSPDVINGSPISQK-IYIKENERFQ 216
Db 568 LIGESTIRCTSDQBERGTWSGPAPLCKLSLLAVQCSHVHIANGYKISKEAPFYFNDIVT 627
QY 217 YKCNWGEYSERGDAVC-TESGWRP-LPSCSEKSCDN----- 251
Db 628 FKCYSGFTLKSGSIQIRCKADNTWDPEIPVCEKTCQHVYRQSLQELPAGSRVELVNTSCQD 687
QY 252 -----PYIPNGDYSPLRIKH-RTGDEITYQ 275
```

```
Db 688 GYQLTGHAYOMCODAENGWFKKIPLCVKIHPHPVIVNGKRTGMMAENFLYGNEVSYE 747
QY 276 CRNGFYPATRGNTAKTST-----GWI-PAPRCTLKP-----CDYPDIKHG 315
Db 748 CDQGGFY-LLGEKKLCGRSDSKHGHSWSPQCLRSPFVTRCPNPEVKHG 796

RESULT 10
CR2_MOUSE
ID CR2_MOUSE STANDARD; PRT; 1025 AA.
AC P19070;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Complement receptor type 2 precursor (Cr2) (Complement C3d receptor).
GN CR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C.
RX MEDLINE=90229735; PubMed=2139457;
RA Fingerth J.D.;
RT "Comparative structure and evolution of murine CR2. The homolog of
RT the human C3d/EBV receptor (CD21).";
RL J. Immunol. 144:3458-3467(1990).
RN [2]
RP SEQUENCE OF 12-1025 FROM N.A.
RX MEDLINE=91010789; PubMed=2145366;
RA Molina H., Kinoshita T., Inoue K., Carel J.C., Holers V.M.;
RT "A molecular and immunochemical characterization of mouse CR2.
RT Evidence for a single gene model of mouse complement receptors 1 and
RT 2.";
RL J. Immunol. 145:2974-2983(1990).
RN [3]
RP SEQUENCE OF 343-401 AND 991-1025 FROM N.A.
RX MEDLINE=89098890; PubMed=2783485;
RA Fingerth J.D., Benedict M.A., Levy D.N., Strominger J.L.;
RT "Identification of murine complement receptor type 2.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:242-246(1989).
RN [4]
RP SEQUENCE OF 289-1025 FROM N.A.
RX MEDLINE=89381350; PubMed=2528587;
RA Kurtz C.B., Paul M.S., Aegerter M., Weis J.J., Weis J.H.;
RT "Murine complement receptor gene family. II. Identification and
RT characterization of the murine homolog (Cr2) to human CR2 and its
RT molecular linkage to Cr2.";
RL J. Immunol. 143:2058-2067(1989).
CC -!- FUNCTION: RECEPTOR FOR COMPLEMENT C3D. PARTICIPATES IN B
CC LYMPHOCYTES ACTIVATION.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: B LYMPHOCYTES.
CC -!- SIMILARITY: TO HUMAN C3D/EBV RECEPTOR (CD21).
CC -!- SIMILARITY: CONTAINS 15 SUSHI (SCR) DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M81083; AAA37451.1; -
CC EMBL; M35684; AAA37448.1; -
CC EMBL; M61132; AAA63295.1; -
CC EMBL; M35685; AAA37450.1; ALT_SEQ.
CC PIR; A43526; A43526.
CC HSSP; P10998; 1LVD.
```



```
FT DOMAIN 453 516 SUSHI 8.
FT DOMAIN 523 579 SUSHI 9.
FT DOMAIN 581 647 SUSHI 10.
FT DISULFID 25 76 BY SIMILARITY.
FT DISULFID 59 87 BY SIMILARITY.
FT DISULFID 91 135 BY SIMILARITY.
FT DISULFID 118 146 BY SIMILARITY.
FT DISULFID 153 197 BY SIMILARITY.
FT DISULFID 180 208 BY SIMILARITY.
FT DISULFID 213 255 BY SIMILARITY.
FT DISULFID 241 267 BY SIMILARITY.
FT DISULFID 274 316 BY SIMILARITY.
FT DISULFID 302 327 BY SIMILARITY.
FT DISULFID 336 378 BY SIMILARITY.
FT DISULFID 364 389 BY SIMILARITY.
FT DISULFID 396 439 BY SIMILARITY.
FT DISULFID 425 450 BY SIMILARITY.
FT DISULFID 454 505 BY SIMILARITY.
FT DISULFID 486 515 BY SIMILARITY.
FT DISULFID 524 567 BY SIMILARITY.
FT DISULFID 553 578 BY SIMILARITY.
FT DISULFID 582 636 BY SIMILARITY.
FT DISULFID 616 646 BY SIMILARITY.
FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 545 545 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 668 AA; 76078 MW; 80BC9E00A9E53FA6 CRC64;

Query Match 15.78; Score 294; DB 1; Length 668;
Best Local Similarity 28.18; Pred. No. 6.8e-16;
Matches 93; Conservative 41; Mismatches 141; Indels 56; Gaps 21;

QY 13 ELLIGSWS--DQYPEGTAQYKCPGY-----RSLGNVIVCRKGEVVALNPLRKCOKR 65
Db 157 ELEHGYNTORTFKVDIVATCTAGYTTTGKOTGEA--EQQANGW-SLTP--QCNKL 211
QY 66 PCG-----HPGDPFGFTLTGNGVFEYGVKAVYTCNEGYOLLGEINRYCDDTG 115
Db 212 MCSLLRLIENGYFHPVK-----QTYEGDLVFFCHENYLSGS-DLIQCYNFG 259
QY 116 WTNDIPICE--VVKCLPVTAPENKIVSSAMEP-DREYHFGQAVRFVCSNGYKIEGDREM 172
Db 260 WYPSPICEGRNRNCPPPVPLNSKI-----QPHSTTYRHGERVHIECELNFVIQGSSEL 314
QY 173 HCSDDGFSKPKCQVE---ISCKS-PDVINGSPISQIIYKENERFYKCNMGYEISE 227
Db 315 LC-ENGKW-T-EPPKCIIEKEKVAQPPSPVNGVAHPHSEIYSGDKVYRCGGGYSLRG 372
QY 228 RGDVAVCTESGRWPLPSCBE--KSC-DNPYIPNG-DYSLPLRIKHRTGDEITYQCRNGFYPA 283
Db 373 SSTITCNRGRWTLPEECVENIENCKPPDPIANGVVVDGLASYTTGSSVEYRC-NEYVLL 431
QY 284 TRGNTAKTSTGWIPAPRCTLKPDYDIKH 314
Db 432 KGSETSRCEQAGWSSPPVC-LEPCTI-DVDH 460

RESULT 14
LEM3_HUMAN
ID LEM3_HUMAN STANDARD; PRT; 830 AA.
AC P16109;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE P-selectin precursor (Granule membrane protein 140) (GMP-140) (PDGEM)
DE (CD62P) (leukocyte-endothelial cell adhesion molecule 3) (LECAM3).
GN SELP OR GMPR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE-89168432; PubMed-2466574;
RA Johnston G.I., Cook R.G., McEver R.P.;
RT "Cloning of GMP-140, a granule membrane protein of platelets and
endothelium: sequence similarity to proteins involved in cell
adhesion and inflammation.";
RL Cell 56:1033-1044(1989).
RN [2]
RP PALMITOYLATION.
RX MEDLINE-93266599; PubMed-7684381;
RA Fujimoto T., Stroud E., Whalley R.E., Prescott S.M., Muszbek L.,
Laposata M., McEver R.P.;
RT "P-selectin is acylated with palmitic acid and stearic acid at
cysteine 766 through a thioester linkage.";
RL J. Biol. Chem. 268:11394-11400(1993).
RN [3]
RP STRUCTURE BY NMR OF 160-199.
RX MEDLINE-97057176; PubMed-8901515;
RA Freedman S.J., Sanford D.G., Bachovchin W.W., Furie B.C., Baleja J.D.,
Furie B.;
RT "Structure and function of the epidermal growth factor domain of P-
selectin.";
RL Biochemistry 35:13733-13744(1996).
RN [4]
RP 3D-STRUCTURE MODELING OF 42-161.
RX MEDLINE-94093388; PubMed-7505880;
RA Bajorath J., Stenkamp R., Aruffo A.;
RT "Knowledge-based model building of proteins: concepts and examples.";
RL Protein Sci. 2:1798-1810(1993).
RN [5]
RP VARIANTS ASN-331; ASP-603; VAL-640 AND PRO-756.
RX MEDLINE-983334547; PubMed-9668170;
RA Heurmann S.M., Ricard S., Nicaud V., Mallet C., Evans A.,
Ruidavets J.B., Arveller D., Luc G., Cambien F.;
RT "The P-selectin gene is highly polymorphic: reduced frequency of the
Pro715 allele carriers in patients with myocardial infarction.";
RL Hum. Mol. Genet. 7:1277-1284(1998).
CC -1- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: STORED IN THE ALPHA-GRANULES OF PLATELETS
AND WEIBEL-PALADE BODIES OF ENDOTHELIAL CELLS. UPON CELL
ACTIVATION BY AGONISTS, P-SELECTIN IS TRANSPORTED RAPIDLY TO
THE CELL SURFACE.
CC -1- SIMILARITY: TO OTHER SELECTINS/LECAMs.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 9 SUSHI (SCR) DOMAINS.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD62P entry;
WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd62p.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M60234; AAA35910.1; JOINED.
CC EMBL; M60217; AAA35910.1; JOINED.
CC EMBL; M60218; AAA35910.1; JOINED.
CC EMBL; M60219; AAA35910.1; JOINED.
CC EMBL; M60222; AAA35910.1; JOINED.
CC EMBL; M60223; AAA35910.1; JOINED.
CC EMBL; M60224; AAA35910.1; JOINED.
CC EMBL; M60225; AAA35910.1; JOINED.
CC EMBL; M60226; AAA35910.1; JOINED.
CC EMBL; M60227; AAA35910.1; JOINED.
CC EMBL; M60228; AAA35910.1; JOINED.
CC EMBL; M60229; AAA35910.1; JOINED.
CC EMBL; M60231; AAA35910.1; JOINED.
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:11:38 ; Search time 67.22 Seconds
(without alignments)
846.702 Million cell updates/sec

Title: US-09-316-163-11
Perfect score: 1876
Sequence: 1 EDCNELPPRRNTEILTGSWS.....POIKHGGLYHENMRRPYFPV 329

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1876	100.0	1172	4 Q9NU87	Q9nu87 homo sapien
2	1366	72.8	1236	11 Q91YB6	Q91yb6 rattus norv
3	903	48.1	669	6 Q28085	Q28085 bos taurus
4	486	25.9	1053	13 Q91275	Q91275 parabrax
5	377	20.1	522	6 Q28769	Q28769 papio cynoc
6	356.5	19.0	3567	11 Q9ES77	Q9es77 mus musculu
7	353	18.8	679	11 Q9S254	Q9s254 mus musculu
8	351	18.7	1911	6 Q29528	Q29528 papio hamad
9	350.5	18.7	559	4 Q9UQV2	Q9uqv2 homo sapien
10	350.5	18.7	2039	4 Q16745	Q16745 homo sapien
11	350.5	18.7	2489	4 Q16744	Q16744 homo sapien
12	349.5	18.6	645	12 Q9WRU2	Q9wru2 macaca mula
13	348	18.6	2014	6 Q29530	Q29530 pan troglod
14	342.5	18.3	661	6 Q29531	Q29531 pan troglod
15	339	18.1	522	11 Q28797	Q28797 pan troglod
16	337.5	18.0	560	5 Q22328	Q22328 caenorhabdi

17	335.5	17.9	555	11 Q99JA1	Q99ja1 cavia porce
18	335	17.9	360	12 Q9YT08	Q9ytq8 ateline her
19	333.5	17.8	315	6 Q28770	Q28770 papio cynoc
20	333	17.8	3508	4 Q96RM4	Q96rm4 homo sapien
21	328	17.5	550	12 P88903	P88903 kaposi's sa
22	328	17.5	550	12 Q40912	Q40912 kaposi's sa
23	327.5	17.5	451	13 Q9DEG0	Q9deg0 gallus gall
24	327	17.4	3389	4 Q96Q09	Q96qu9 homo sapien
25	327	17.4	3564	11 Q92313	Q92313 mus musculu
26	323	17.2	497	11 Q63612	Q63612 rattus norv
27	323	17.2	559	11 Q63135	Q63135 rattus norv
28	317	16.9	452	11 Q61407	Q61407 mus musculu
29	315	16.8	481	4 Q9H284	Q9h284 homo sapien
30	311.5	16.6	469	11 Q91X48	Q91x48 mus musculu
31	310	16.5	259	12 P87616	P87616 cowpox viru
32	308.5	16.4	395	12 Q9J2M6	Q9j2m6 macaca mula
33	308.5	16.4	1032	4 Q13866	Q13866 homo sapien
34	306.5	16.3	417	11 Q35520	Q35520 rattus norv
35	306.5	16.3	946	4 Q96Q03	Q96q03 homo sapien
36	306	16.3	303	11 Q61405	Q61405 mus musculu
37	304	16.2	483	11 Q64735	Q64735 mus musculu
38	303.5	16.2	1032	11 Q9DC83	Q9dc83 mus musculu
39	302.5	16.1	657	4 Q14006	Q14006 homo sapien
40	302	16.1	533	11 Q08569	Q08569 cavia porce
41	298	15.9	1124	5 Q9VYR4	Q9vyr4 drosophila
42	294	15.7	1653	5 Q9VIU9	Q9viu9 drosophila
43	293.5	15.6	601	11 Q9CUT3	Q9cut3 mus musculu
44	293.5	15.6	669	11 Q922H0	Q922h0 mus musculu
45	293	15.6	363	6 Q02839	Q02839 sus scrofa

ALIGNMENTS

RESULT 1

Q9NU87					
ID	Q9NU87	PRELIMINARY;	PRT:	1172 AA.	
AC	Q9NU87;				
DT	01-OCT-2000 (TReMBLrel. 15, Created)				
DT	01-OCT-2000 (TReMBLrel. 15, Last sequence update)				
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)				
DE	DJ177P10.1.1 (H FACTOR 1 (COMPLEMENT) ISOFORM 1).				
.N	HFI.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Bird C.;				
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AL049744; CAB70597.1; .				
DR	HSSP; P08603; 1HFH.				
DR	InterPro; IPR000436; Sushi_SCR_CCP.				
DR	Pfam; PF00084; sushi; 19.				
DR	SMART; SM00032; CCP; 19.				
SQ	SEQUENCE 1172 AA; 132087 MW; 8F5B954C4B4FA54 CRC64;				

Query Match 100.0%; Score 1876; DB 4; Length 1172;
Best Local Similarity 100.0%; Pred. No. 1.4e-165;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	EDCNELPPRRNTEILTGSWS	QTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWALNPLR	60
Db	19	EDCNELPPRRNTEILTGSWS	QTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWALNPLR	78
Qy	61	KCQKRPCGHPGDTDFGTFTLTGGNVFEYGVKAVYTCNEGYQLLGEINRYRECDT	DGWTNDI	120
Db	79	KCQKRPCGHPGDTDFGTFTLTGGNVFEYGVKAVYTCNEGYQLLGEINRYRECDT	DGWTNDI	138
Qy	121	PICEVVKCLPVTAPENKGIYSSANEPREYHFGQAVFCVNSGKIEGDEMHSCDDGFW	180	

```
Db 139 PICEVVKLPVTAPENGKIVSSAMEPDREYHFGQAVRFVNCNSGYKIEGDEMHCSDDGF 198
QY 181 SKEKPKCIVEISCKSPDVINGSPISQIIYKENERFOYKCNMGVEYSERGDAVCTESGWRP 240
Db 199 SKEKPKCIVEISCKSPDVINGSPISQIIYKENERFOYKCNMGVEYSERGDAVCTESGWRP 258
QY 241 LPSCEKSCDNPYPNGDYSPLRIKHRTGDEITYQCRNGFYPATRGNTAKCTSTGWIAP 300
Db 259 LPSCEKSCDNPYPNGDYSPLRIKHRTGDEITYQCRNGFYPATRGNTAKCTSTGWIAP 318
QY 301 RCTLKPCDYPDKHGLYHNNRRPFPV 329
Db 319 RCTLKPCDYPDKHGLYHNNRRPFPV 347

RESULT 2
QY1YB6
ID QY1YB6 PRELIMINARY: PRT: 1236 AA.
AC QY1YB6;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE COMPLEMENT INHIBITORY FACTOR H.
GN FH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=LIVER;
RA Demberg T., Goetze O., Schlaf G.;
RT "Rat complement factor H: molecular cloning, sequencing and expression
in tissues and isolated cells.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; A320522; CAC67513.1; -.
SQ SEQUENCE 1236 AA; 140343 MW; 1AC89FFA28232BEF CRC64;

Query Match 72.8%; Score 1366; DB 11; Length 1236;
Best Local Similarity 71.1%; Pred. No. 3.8e-118;
Matches 234; Conservative 31; Mismatches 64; Indels 0; Gaps 0;

QY 1 EDCNELPPRRNTEILTGSDQTYPEGTQAIYKCRPGYSLGNVIMVCRKGWALNPLR 60
Db 19 EDCKGPPPRENSBILSGNSEQLYSEGTAQYKCRPGYRTLGTVKVKNGWVPSNPSR 78
QY 61 KCOKRCPGHGDPFTFTLTGNNVFEYGVKAVYTCNEGQYLLGEINRECDTGTNDI 120
Db 79 ICKKRCGHGDPFTFGSFLAVGSEFEFGAKVYVTCDEGYQLLGEIDYRECDAGTNDI 138
QY 121 PICEVVKLPVTAPENGKIVSSAMEPDREYHFGQAVRFVNCNSGYKIEGDEMHCSDDGF 180
Db 139 PICEVVKLPVTELENGRIVSGNAEPDQYIYFGQVVFECNSGFKIEGOKEMHCSENGLW 198
QY 181 SKEKPKCIVEISCKSPDVINGSPISQIIYKENERFOYKCNMGVEYSERGDAVCTESGWRP 240
Db 199 SNEKPOCIVEISCLPPRVENGDIYLPVYKENERFOYKCKQGFVYKERGDAVCTGSGWNP 258
QY 241 LPSCEKSCDNPYPNGDYSPLRIKHRTGDEITYQCRNGFYPATRGNTAKCTSTGWIAP 300
Db 259 QPSCEEMTCLTYPNGDIYTPRIKHRIIDDEIRYECKNGLYPATRSPVSKCTITGWIAP 318
QY 301 RCTLKPCDYPDKHGLYHNNRRPFPV 329
Db 319 RCTLKPCDYPDKHGLYHNNRRPFPV 347

RESULT 3
QY2805
ID Q2805 PRELIMINARY: PRT: 669 AA.
AC Q2805;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
```

```
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CCP MODULES 3-12, WITH PARTS OF CCP 2 AND 13 (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OX Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=96202005; PubMed=8615824;
RA Soames C.J., Day A.J., Sim R.B.;
RT "Prediction from sequence comparisons of residues of factor H involved
in the interaction with complement component C3b.";
RL Biochem. J. 315:523-531(1996).
DR EMBL; X98697; CAA67257.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 11.
DR SMART; SM00032; CCP; 11.
FT NON_TER 1
FT NON_TER 669
SQ SEQUENCE 669 AA; 75683 MW; D0D9DB30EE747AC2 CRC64;

Query Match 48.1%; Score 903; DB 6; Length 669;
Best Local Similarity 60.9%; Pred. No. 1.9e-75;
Matches 154; Conservative 33; Mismatches 64; Indels 2; Gaps 1;

QY 76 GTFTLTGNNVFEYGVKAVYTCNEGQYLLGEINRECDTGTNDIPTCEVVKLPVTAPE 135
Db 3 GSPHLAEGNQFEYGAHVYTCDEGYQWGMENFRECDTGTNDIPTCEVVKLPVTEPE 62
QY 136 NKKIVSSAMEPDREYHFGQAVRFVNCNSGYKIEGDEMHCSDDGFVSKERPKCIVEISCKSP 195
Db 63 NGKIFSDALEPDQETTYGQVVFECNSGYMLDGPQIHCAGVWSAETPKCIVEIFCKPP 122
QY 196 DVINGSPISQIIYKENERFOYKCNMGVEYSERGDAVCTESGWRPLPSCPEKSCDNPYP 255
Db 123 VILNGQAVLPKATYKQNERVQYRCAAGFEYQGQGDIVCTKSGWTAPTCLIEITCDPPRIP 182
QY 256 NGDYSPLRIKHRTGDEITYQCRNGFYPATRGNTAKCTSTGWIAPRCTLKPCDYPDIKHG 315
Db 183 NGVYRPELSKYRQDKITVECKKGFFPEIRGTATCTRDGQVVPVPCAMKPCSYPIKHG 242
QY 316 GLYHNNRRPFPV 328
Db 243 RLYYS--YRGYFP 253

RESULT 4
QY1275
ID QY1275 PRELIMINARY: PRT: 1053 AA.
AC QY1275;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE COMPLEMENT REGULATORY PLASMA PROTEIN.
OS Parabrax nebulifer (barred sand bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Serranidae; Paralabrax.
OX NCBI_TaxID=30873;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=94318039; PubMed=8042982;
RA Dahmen A., Kaidoh T., Zipfel P.F., Gigli I.;
RT "Cloning and characterization of a cDNA representing a putative
complement-regulatory plasma protein from barred sand bass (Parabrax
nebulifer).";
```

```
RL Biochem. J. 301:391-397(1994).
DR EMBL; L21703; AAA92556.1; -.
DR HSSP; P08603; LHFH.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 16.
DR SMART; SM00032; CCP; 16.
SQ SEQUENCE 1053 AA; 117597 MW; F27E32C3AD76D5D3 CRC64;

Query Match 25.9%; Score 486; DB 13; Length 1053;
Best Local Similarity 29.7%; Pred. No. 1.8e-36;
Matches 114; Conservative 48; Mismatches 134; Indels 88; Gaps 13;

QY 21 DQTPPECTQAIYKCRGYSRLGNVIMVCRKGEWVALNPLRKKCKRCGHPGDPFTFTL 80
DB 47 EASYPGGROVRVGCNVGYS--GFFKLVCVEGKWETRG--AKQPRSCGHPGDAQOAFHL 102
QY 81 TCGNVEYGVKAVYTCNEGQYLLGEINRECDTGDWTNDIPICEVVKCLPVTAPENGKIV 140
DB 103 AEGNDFVFGSKVYTCQKGQWYSRINRYRCVAEGHDGVVPCESQOCPLIHVDNNOVVI 162
QY 141 SSAMEPDREYHFGQAVRVCNYSYKI--EGDEEMHCSDDGFWSEKPKVCVEISCKSPDVIN 199
DB 163 GG----PEEATFCNVVRFSCKSRSEILDGSPELYCDERGDWSGVPVKRAITCAIPIEN 218
QY 200 GSPISQKIYKENERQYKCNMGYEYSERGDAVCTESG-----WRPLPSCEEKSC 249
DB 219 GNVPGAIREYKENDVLHYECDRAFKHIDR--PSTCIKQIGKAWSPTPLCESIKRLTIMD 277
QY 250 -----DNPIYIP-----NGDYS----- 260
DB 278 GTRYEPAYRNLFSPGETLVKICARTSWISTPQTSVVTTCODNGEWSIRPTQOEVRCSNR 337
QY 261 -PLRI-----KHTGDEITYOCNRYFPATRGNT-AKCTSTGWIPAPRCTLKPCDY 309
DB 338 RPEHVDSDVRSWERYTLDNTRYCKRG-YKRTGGVWATCGRNGMWPPLCEVKTCSK 396
QY 310 PDIKHG-----GLYHENWRRPY 326
DB 397 ENIQDAVIVGTDKQIYNLNOKAIY 420

RESULT 5
Q28769 ID Q28769 PRELIMINARY; PRT; 522 AA.
AC Q28769
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE COMPLEMENT RECEPTOR (FRAGMENT).
OS Papio cynocephalus (Yellow baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Papio.
OX NCBI_TaxID=9556;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RA Birmingham D.J., Logar C.M., Shen X.-P., Chen W.;
RT "The baboon erythrocyte complement receptor is a glycoprophosphatidyl
RT inositol-linked protein encoded by a homologue of the human CRI-like
RT genetic element."
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L77977; AAA99004.1; -.
DR HSSP; F10998; 1VVD.
DR InterPro; IPR002396; Selectin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 7.
DR PRINTS; PR00343; SELECTIN.
DR SMART; SM00032; CCP; 7.
KW Receptor.
FT NON_TER
SQ SEQUENCE 522 AA; 56626 MW; 312FCBE03ADF19DC CRC64;
```

```
Query Match 20.1%; Score 377; DB 6; Length 522;
Best Local Similarity 30.2%; Pred. No. 1e-26;
Matches 107; Conservative 55; Mismatches 142; Indels 50; Gaps 22;

QY 3 CN---ELPPRRNTEILTGSDQTYPEGTQAIYKCRGYSRLGNVIMVCRKGEWVALNPL 59
DB 48 CNAPEQLPFARPTNLTDAS----EFVGYTLKYECLEPGYHGKPFIIICLKNVWTSARD- 102
QY 60 RKQKRCPCGHPGDPFTFTLTGNNVFEYGVKAVYTCNEGQYLLGEINRECDTGDG--- 115
DB 103 -KTRKSCRNPDPVNGWHVI--KDIQFGSQINYSCKNGIRLIGS--SSATCIISGTVI 158
QY 116 WTNDIPICEYVKC-LPVTAPENKIVSSAMEPDREYH-FGQAVRFVCSNG-----YKIE 167
DB 159 WDNETPICEIIPGGLPPTI-ANGDFIST-----REYFPGYGVVYTRCNLGRKKLFELV 213
QY 168 GDEMHCS--DD--GFWSEKPKVC-VEISCKSPDVINGSPIS-OKIYKENERFOYKCNM 221
DB 214 GEPISYCTSKDDQVGWISGPAQCIIPNKCMPNVENGVLVSVNRSLSLNEVVEFRCP 273
QY 222 GYEYSERGDVCTE-SGWRP-LPSCEEKSCDNPYPNGDYSPLRKHRTGDEITYOCNNG 279
DB 274 GFYMKGPRHVOCALNKWPELPSCSRVCQPPPEILLGHEHTPSHQDFSPGOEYFVSCPG 333
QY 280 FYPATRGNTAKCTSTG-WIP-APRCTLKPCD--YPODIKHGGLYHENRRRPFVFPV 329
DB 334 -YDLRGAASLHCTPQGDWNPPEAPICIVKSCDDELGLQLPHG-----RVLEPL 378

RESULT 6
Q28577 ID Q28577 PRELIMINARY; PRT; 3567 AA.
AC Q28577
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE POLYDOM PROTEIN PRECURSOR.
GN POLYDOM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/HENSIC.
RX MEDLINE=20517255; PubMed=11062057;
RA Gilges D., Vinit M.-A., Callebaut I., Coulombel L., Cacheux V.,
RA Romeo P.-H., Vigon I.;
RT "Polydom : a secreted protein with pentraxin, complement control
RT protein, epidermal growth factor and von willebrand factor A
RT domains."
RL Biochem. J. 352:49-59(2000).
DR EMBL; AF206329; AAG32160.1; -.
DR HSSP; P00740; IEDM.
DR MGD; MGI:1928849; Polydom.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR003410; HYR.
DR InterPro; IPR001759; Pentaxin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00008; EGF; 10.
DR Pfam; PF02494; HYR; 2.
DR Pfam; PF00084; sushi; 33.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR000895; PENTAXIN.
DR PRINTS; PR00453; VWFADOMAIN.
```

DR	PTDOM; PD002153; Pentaxin; 1.
DR	SMART; SM00032; CCP; 34.
DR	SMART; SM00181; EGF; 15.
DR	SMART; SM00179; EGF_CA; 9.
DR	SMART; SM00001; EGF_like; 3.
DR	SMART; SM00159; PTX; 1.
DR	SMART; SM00327; VWA; 1.
DR	PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR	PROSITE; PS02234; VWFA; 1.
KW	Signal.
FT	SIGNAL
SQ	SEQUENCE 3567 AA; 387391 MW; 8FBA8276E12293E5 CRC64;
Query, Match	19 0%; Score 356.5; DB 11; Length 3567;
Best Local Similarity	29.2%; Pred. No. 9e-24;
Matches	90; Conservative 39; Mismatches 146; Indels 33; Gaps 13;
Qy	17 GSWSDDTYPECTAIYKCRPGYRSLGNVIMVC-RKGWVALNPLRKCKOKRCPGHPGDTPF 75
Db	1797 GHSSGEIYTCVTATFSFCDEGHVLGVSTITCLETGWDLRP--SCAISCGRP---PV 1851
Qy	76 GTFLTGCNVFEYGVKAVYTCNEYQLLGELINRECDTGTNDIPICEVVKLCVPVTAPE 135
Db	1852 PENGVGDSAGTYGSVVYRGDKYTILSGDESACLASGWSHSPVELVKCSQPEDIN 1911
Qy	136 NGKITVSAMEPDREHYFHQAARFFVCNSGIYLEGDEEHMCHSDDGFWSEKPKPCVEISCKSP 195
Db	1912 NGKIYLSGLT-----YLSTASYSCENGYSLQGPSLLLECTASGWDRAAPPSCQLVSCGEP 1965
Qy	196 DVINGSPIISKIIYKENRF-----QYKNMGYEYSERGDVAICTESG-WRPL-PSCEEKSC 249
Db	1966 PIVDA-----VIRGSNFTFGNTVAYTKEGYTLAGPDTIQCAQNGKWNNSHQCCLAVSC 2020
Qy	250 DNPYPNGDYSPLRKHRT-GDEITYOCRNIFYPATRCNTAKCTSTG-WIP-----APRC 302
Db	2021 DEP--PNVDHRASPETAHLRFLGDTAFYYCADG-YSLADNSQLICNAQQGNWVPVPAQAVPRC 2077
Qy	303 TLKPCCDPY 310
Db	2078 IAHFCEKP 2085
RESULT 7	
Q99254	
ID	Q99254 PRELIMINARY; PRT; 679 AA.
AC	Q99254;
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	COMPLETE RECEPTOR TYPE 2 (CR2) (FRAGMENT).
GN	CR2.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID	10090;
RN	[1]
RC	SEQUENCE FROM N.A.
RC	STRAIN=BALB/C;
RA	MEDLINE=90229754; PubMed=2139460;
RA	Kurtz C.B., O'Toole E., Christensen S.M., Wells J.H.;
RT	"The murine complement receptor gene family. IV. Alternative splicing
RT	of Cr2 gene transcripts predicts two distinct gene products that share
RT	homologous domains with both human CR2 and CR1.";
RL	J. Immunol. 144:3561-3591(1990).
RN	[2]
RP	SEQUENCE OF 21-367 FROM N.A.
RX	MEDLINE=95105691; PubMed=7528766;
RA	Kim Y.U., Kinoshita T., Molina H., Hourcade D., Seya T., Wagner L.M.,
RA	Holers V.M.;
RT	*Mouse complement regulatory protein Crry/p65 uses the specific
RT	mechanisms of both human decay-accelerating factor and membrane
RT	cofactor protein.";


```
DR Pfam; PF00084; sushi; 29.
DR SMART; SM00032; CCP; 29.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_4.
DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW Receptor.
FT NON_TER. 1911 1911
SQ SEQUENCE 1911 AA; 210174 MW; 535A4DD0EAF521D CRC64;

Query Match      18.7%; Score 351; DB 6; Length 1911;
Best Local Similarity 29.0%; Pred. No. 1.4e-23;
Matches 93; Conservative 57; Mismatches 135; Indels 36; Gaps 21;

QY 20 SDQTYPEGTQAIYKCRGYSRISGNIMV-CRKG-EWVALNLRKCKRKGHPGDTPEGT 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1299 NDFEFPFVSLNYECPGV--PGRMFSISCLNVLWSSVED--NCRKRSCTGTPPE-PFNG 1353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 78 FTLTGGNVFEYGVKAVYTCNCGYOLLGEINRECDTGG----WTNDIPICEVVKLPVTA 133
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1354 MYHINTDT-OFGSTVNYSCNCGYOLLGEINRECDTGG----WTNDIPICEVVKLPVTA 1411
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 134 PENGKIVSSAMPEPDREYHFGQAVRFVCSNG-----YKIEGDEEMHCS--DD--GFWSKE 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1412 ISNGDFYSNNR--TSFHSQVTVTYOCHTGPDPGEQLFELVGRSIIYCTSKDOVGAWSSP 1468
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 184 KFKCVKELIS-CKSPDVINGSPI-SQKIIYKENERFOYKCMGVEYSEKDAVC-TESGNRP 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1469 PPRCISTNKCTAPEKNGIRVPGNRFSFTSLNIVRFRCQPGFVMVGSHTVQCQTNNRNGP 1528
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 241 -LPSCEEKSCNDPIYPNGDYSPL-RIKHRTGDEITYQCRNGFYPATRGNTAKCTSTG-WI 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1529 KLPHCSRVCOPPELHGEHTPSHQDKSPGQEVFYSCEPG-YDLRGAASLHCTFGQDWS 1587
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 298 P-APRCTLKPCD--YPDIKHG 315
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1588 PEAPICTVKSCDFELGQLPHG 1608
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
QSUQV2 PRELIMINARY; PRT; 559 AA.
ID Q9UQV2 AC Q9UQV2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CR1 C3B/C4B RECEPTOR SECRETED FORM PRECURSOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89010527; PubMed=2971757;
RA Hourcade D., Miesner D.R., Atkinson J.P., Holers V.M.;
RT "Identification of an alternative polyadenylation site in the human
RT C3b/C4b receptor (complement receptor type 1) transcriptional unit and
RT prediction of a secreted form of complement receptor type 1.";
RL J. Exp. Med. 168:1255-1270(1988).
DR EMBL; X14362; CAA32541.1; JOINED.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR002396; Selectin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR000834; Zn_carbopept.
DR Pfam; PF00084; sushi; 8.
DR PRINTS; PRO0343; SELECTIN.
DR SMART; SM00032; CCP; 8.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
KW Signal; Receptor.
FT NON_TER. <1 1
FT SIGNAL 17 559 POTENTIAL.
FT CHAIN 17 559 POTENTIAL.
SQ SEQUENCE 559 AA; 61424 MW; DBFFE965CA179D75 CRC64;

Query Match      18.7%; Score 350.5; DB 4; Length 559;
Best Local Similarity 29.3%; Pred. No. 3.3e-24;
Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;

QY 3 CNE---LPPRRNTEILTGSWSQDTYPEGTQAIYKCRGYSRISGNIMVCRKKGWVALNPL 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 18 CNAPEWLFPARPTNLT---DEFEFPDGYLNYECPGYSGRPFSTICLKNSVMTGAKD- 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 RKCORPCGHPGDTPEGTFTLTGGNVFEYGVKAVYTCNCGYOLLGEINRECDTGG---DTDG 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 73 -RCRRKSCRNPDPVNGMVHVIKNG---IQFGSIKYSCTKGYRLIGS-SSATCIISGDTVI 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 WTNDIPICSVKRC-LPVTAPENGKIVSSAMPEPDREYHFGQAVRFVCSNG-----YKIEG 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 129 WNEPPICDIRICGLPPTI-TNGDFISTNRE---NFHYGSVTVYRCNPGSGGRKVFELVG 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 169 DEMHC--SDD--GFWSKEKPKC-VEISKSPDVINGSPI-SQKIIYKENERFOYKCMNG 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 185 EPSIYCTSDNDQGVGWSGPAQCIIIPNKCTPPNVENGILVSDNRSLSFSLNEVVFRCQPG 244
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 223 VEYSEKGDVCTE-SQWRP-LPSCEEKSCNDPIYPN-----GDYSPLRKIHRTG 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 245 FVWKGPRRVKCOALNKWPELPSC-SRVQPP--PDVLAERTQRDKDNFSP-----G 294
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 270 DEITYQCRNGFYPATRGNTAKCTSTG-WIP-APRCTLKPCD 308
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 QEVFYSCEPG-YDLRGAASMRCTPGQDWSPPAAPTCEVKSCD 334
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
Q16745 PRELIMINARY; PRT; 2039 AA.
ID Q16745 AC Q16745;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE COMPLEMENT RECEPTOR 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94065175; PubMed=8245463;
RA Vik D.P., Wong W.W.;
RT "Structure of the gene for the F allele of complement receptor type 1
RT and sequence of the coding region unique to the S allele.";
RL J. Immunol. 151:6214-6224(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Vik D.P., Wong W.W.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L17418; AAB60694.1; JOINED.
DR EMBL; L17390; AAB60694.1; JOINED.
DR EMBL; L17399; AAB60694.1; JOINED.
DR EMBL; L17409; AAB60694.1; JOINED.
DR EMBL; L17419; AAB60694.1; JOINED.
DR EMBL; L17420; AAB60694.1; JOINED.
DR EMBL; L17421; AAB60694.1; JOINED.
DR EMBL; L17422; AAB60694.1; JOINED.
DR EMBL; L17423; AAB60694.1; JOINED.
DR EMBL; L17391; AAB60694.1; JOINED.
DR EMBL; L17392; AAB60694.1; JOINED.
DR EMBL; L17393; AAB60694.1; JOINED.
DR EMBL; L17394; AAB60694.1; JOINED.
DR EMBL; L17395; AAB60694.1; JOINED.
DR EMBL; L17396; AAB60694.1; JOINED.
DR EMBL; L17397; AAB60694.1; JOINED.
DR EMBL; L17398; AAB60694.1; JOINED.
DR EMBL; L17400; AAB60694.1; JOINED.
DR EMBL; L17401; AAB60694.1; JOINED.
DR EMBL; L17402; AAB60694.1; JOINED.
```



```
QY 169 DEEMHC--SDD--GFWSKPKKC-VEISCKSPDVINGSPIS-OKIYKENERFOYKCNMG 222
Db 210 EFSIYCTSDNDQGVIGWSPAPQCIIPNKCITPPNVENGLVSDNRSLFSLNVEVFRCPQG 269
QY 223 YEYSEGDVCTE-SWRP-LPSCCEKSCDNYIIN-----GDYSLRKHRTG 269
Db 270 FVWKGPRRVKCOALNKWPELPSC-SRVCOPP--PDVLHAERTQDKDNFSP-----G 319
QY 270 DEITYQCRNGFYPATRGNTAKTSTG-WIP-APRCTLKPCD 308
Db 320 QEVFYSCEPG-YDLRGAASMRCTPGQDWSPAAPTCEVRSCD 359

RESULT 12
Q9WRU2
ID Q9WRU2 PRELIMINARY; PRT: 645 AA.
AC Q9WRU2;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE COMPLEMENT BINDING PROTEIN.
OS Macaca mulatta rhadinovirus 17577.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=83534;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99174001; PubMed=10074154;
RA Searles R.P., Bergquam E.P., Axthelm M.K., Wong S.W.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF083501; AAD21332.1;
DR HSSP: P10998; 1VVD.
DR InterPro: IPR000436; SUSHI_SCR_CCP.
DR Pfam: PF00084; sush1; 8.
DR SMART: SM00032; CCP; 8.
SQ SEQUENCE 645 AA; 71526 MW; 93D8DE35ABF61EB2 CRC64;

Query Match 18.6%; Score 349.5; DB 12; Length 645;
Best Local Similarity 25.8%; Pred. No. 4.9e-24;
Matches 99; Conservative 45; Mismatches 164; Indels 75; Gaps 14;

QY 1 EDCNELPPRRNTEILTGSMSDQTYPEGTQAIYKCRPGYRSLGNVIMV--CRKGEWALNP 58
Db 21 ENCK--PPHFTYRKYKSTKDLISVGETAELICRPGYVTVTKIITTECLONGTW--STP 76
QY 59 LRKQKRCPCGHPGDPFTFTLTG-VNFEYGVKAVYTCNEGYYQLLGEINREC---DTD 114
Db 77 NFPCDRKRCPPTADLLNGAVIHGGDNALKFGSNISYECNEGYDLIGS-NVRFCLQDTE 135
QY 115 --GWTNDIPICEVVKCLPVTAPENKIVSSAMEPDRE-YHFGQAVRVCNSGYKIEGDEE 171
Db 136 NVNWDNSPEVCEIOKCIKPPAVEHGDVL-----PNQDVYNGDAITFKCSLSTLVGSTT 190
QY 172 MHCSDDCGFWSKEKPKVCISCKSPDVING-SPISQKIYKENERFOYKCNMGYSEKGD 230
Db 191 LVCTSNKWSNSFTCLMLVCESPOIDNGYIDIGLSRRYNHGQSITVKCSDGYNVGPET 250
QY 231 AVCTESGW-RPLPSC-----EEKSCDNPYIPNGDYSPLRI----- 264
Db 251 LTCNTTWTVPPLPKCVLVTNNPSTPMPETPTPDYQKINLSTAKTATTPNAFTVTV 310
QY 265 -----KHRTGDEITYQCRNGFYPATRGNTAKTSTGWI 297
Db 311 VSPEKDDVTCVPHFRFMVKAENDKEKYSVGASVELICRPGFTKMQSTVSVECLSNGTW 370
```

```
QY 298 PAP--RCTLKPCDYPDIDIKHGGLY 318
Db 371 TAPNAKCHRRKCKPTPOELLNGEY 393

RESULT 13
Q29530
ID Q29530 PRELIMINARY; PRT: 2014 AA.
AC Q29530;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE COMPLEMENT RECEPTOR 1 (FRAGMENT).
GN CRI.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94292799; PubMed=8021505;
RA Birmingham D.J., Shen X.P., Hourcade D., Nickells M.W., Atkinson J.P.;
RT "Primary sequence of an alternatively spliced form of CRI. Candidate
RT for the 75,000 M(r) complement receptor expressed on chimpanzee
RT erythrocytes."
RL J. Immunol. 153:691-700(1994).
DR EMBL: L24920; AAA51438.1;
DR HSSP: P08603; 1HEI.
DR InterPro: IPR001424; SOD_CU_ZN.
DR InterPro: IPR000436; SUSHI_SCR_CCP.
DR InterPro: IPR000834; Zn_carboOpept.
DR Pfam: PF00084; sush1; 30.
DR SMART: SM00032; CCP; 30.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR PROSITE: PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW Receptor.
FT NON_TER 1 2014
FT NON_TER 2014 2014
SQ SEQUENCE 2014 AA; 221281 MW; 6D6C3A74D81F1DB9 CRC64;

Query Match 18.6%; Score 348; DB 6; Length 2014;
Best Local Similarity 28.4%; Pred. No. 2.7e-23;
Matches 96; Conservative 57; Mismatches 135; Indels 50; Gaps 23;

QY 20 SDQTYPEGTQAIYKCRPGYRSLGNVIMV-CRKG-EMVALNPLRKCKRCPGHGDPFGT 77
Db 1387 NDFEPVGTSLNYECPGT--FGKMFSSCLLENLVWSSVED--NCRKSCGPPPE-PFNG 1441
QY 78 FTLTGNNVFEYGVKAVYTCNEGYYQLLGEINRECDDTG----WTNDIPICEVVKCLPVTA 133
Db 1442 MVHINTDT-QFGSTVYSCNEGFRILIGSPS-TTCLVSGNNVTMDKKAPICEIICEPPT 1499
QY 134 PENGKIVSSAMEPDREYHFGQAVRVCNSG-----YKIEGDEMHCS--DD--GFWSKE 183
Db 1500 ISNGDFYSNNR---ASFHNGTVTVYQCHTGPDEQLFELVGRSICYTSKDDQGVWSSP 1556
QY 184 KPKCVEIS-CKSPDVINGSPI-SQKIIYKENERFOYKCNMGYSEKGDVCTESG-WRP 240
Db 1557 PPRCISTNKTCTAPEVNAIRVPGNRSFSLTEIVFRCPGPFVMSGHVTVCQOTGRMGP 1616
QY 241 -LPSCEEKSCDNPYIPNGDYSPLRIKHR-----TGDEITYQCRNGFYPATRGNTAKTSTG 295
Db 1617 KLPHCSRVCCPPPELHGEHTP---SHQDNFSPGQEVFYSCEPG-YDLRGAASLHCTQGG 1672
QY 296 -WIP-APRCTLKPCD--YPDIKHGGLYHENMRRYFPV 329
Db 1673 DWSPEAPRCTVKSCDDFLGQLPHG-----RVLFPL 1702

RESULT 14
Q29531
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:03:37 ; Search time 75.55 Seconds
(without alignments)
483.697 Million cell updates/sec

Title: US-09-316-163-11
Perfect score: 1876
Sequence: 1 EDCNELPPRRNTTEILTGWS.....PDIKHGGLYHNNRRPYFPV 329

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1362	72.6	240	AAW39154	Human partial Comp
2	1229	65.5	216	AAW39155	Clone PR89FH410 C
3	353	18.8	456	AAV55752	Human CR1 protein
4	353	18.8	581	AAAI3490	Human C4 binding p
5	352.5	18.8	543	AAAR28570	CR1-4 (266-274) KIK
6	350.5	18.7	778	AAAR73147	Amino acid sequenc
7	350.5	18.7	1930	AAW45899	Human complement r
8	350.5	18.7	2039	AAV55751	Human C3b/C4b rece
9	350.5	18.7	2039	ABG00287	Novel human diagno
10	350.5	18.7	2044	ABB11782	Human CR1 protein
11	350.5	18.7	2044	AAW39224	Human polypeptide

12	350.5	18.7	2044	22	AAW41010	Human polypeptide
13	350.5	18.7	2317	10	AAAP92219	CR1 protein. Homo
14	349.5	18.6	645	21	AAAB53125	Macaca mulatta rha
15	347.5	18.5	543	13	AAAR28557	CR1-4 (99H, 103E)
16	345.5	18.4	543	13	AAAR28547	CR1-4 (52S, 53S, 5
17	343.5	18.3	543	13	AAAR28560	CR1-4 (114S) analo
18	341.5	18.2	543	13	AAAR28550	CR1-4 (64K) analog
19	341.5	18.2	543	13	AAAR28553	CR1-4 (85R, 87N) a
20	341.5	18.2	543	13	AAAR28565	CR1-4 (121Q) analo
21	341.5	18.2	543	13	AAAR28568	CR1-4 (347T, 349Y)
22	341.5	18.2	543	13	AAAR28569	CR1-4 (369-376 STK
23	341.5	18.2	543	13	AAAR28571	CR1-4 (364-367 NAA
24	341.5	18.2	2039	12	AAAR11810	Human complement t
25	340.5	18.2	543	13	AAAR28545	CR1-4 (37Y) analog
26	340.5	18.2	543	13	AAAR28548	CR1-4 (57V, 59K) a
27	340.5	18.2	1497	22	AAAG93954	Human polypeptide,
28	339.5	18.1	543	13	AAAR28555	CR1-4 (92T) analog
29	339.5	18.1	543	13	AAAR28558	CR1-4 (109N, 110A,
30	339.5	18.1	543	13	AAAR28563	CR1-4 (117P) analo
31	339.5	18.1	543	13	AAAR28567	CR1-4 (318-321 RNP
32	338.5	18.0	543	13	AAAR28549	CR1-4 (64K, 65T) a
33	338.5	18.0	543	13	AAAR28551	CR1-4 (65T) analog
34	338.5	18.0	543	13	AAAR28562	CR1-4 (116K) analo
35	336.5	17.9	543	13	AAAR28556	CR1-4 (94H) analog
36	336.5	17.9	543	13	AAAR28564	CR1-4 (116K, 117P)
37	336.5	17.9	2039	14	AAAR36743	CR1. Homo sapiens
38	336	17.9	1537	12	AAAR11982	Partial human comp
39	336	17.9	1929	22	ABG00103	Novel human diagno
40	335.5	17.9	543	13	AAAR28552	CR1-4 (78T, 79D) a
41	335.5	17.9	543	13	AAAR28561	CR1-4 (115T) analo
42	335.5	17.9	543	13	AAAR28566	CR1-4 (318R, 319N)
43	334.5	17.8	543	13	AAAR28554	CR1-4 (92T, 94H) a
44	333.5	17.8	543	13	AAAR28544	CR1-4 (35E) analog
45	333.5	17.8	543	13	AAAR28546	CR1-4 (44T, 47D, 4

ALIGNMENTS

RESULT 1

AAW39154

ID AAW39154 standard; Protein; 240 AA.

XX

AC AAW39154;

XX

DT 27-APR-1998 (first entry)

XX

DE Human partial Complement factor H protein fragment 1.

XX

DE Complement factor H; tumour associated antigen; renal cancer;

KW urogenital cancer; medicament; modulator.

XX

OS Homo sapiens.

XX

PN WO9738136-A1.

XX

PD 16-OCT-1997.

XX

PF 09-APR-1997; 97WO-US057110.

XX

PR 06-MAR-1997; 97US-0812481.

PR 09-APR-1996; 96US-0015083.

PR 09-APR-1996; 96US-0630048.

PR 06-MAR-1997; 97US-0038614.

XX

PA (BARD-) BARD DIAGNOSTIC SCI INC.

XX

PI Enfield DL, Hass GM, Kinders RJ;

XX

DR WPI; 1997-512742/47.

DR N-PSDB; AAV02790.

XX

PT Treating or screening for cancer, e.g. renal or urogenital cancer -

PT by modulating or detecting tumour associated human complement Factor
 PT H related antigen, or nucleic acid encoding it
 XX
 PS Example 6B; Fig 6B; 104pp; English.
 XX
 CC This partial protein sequence represents a region of the human
 CC tumour-associated complement factor H (CFH). This sequence is used
 CC in the identification of complement factor H related proteins and
 CC antigens isolated from clone PRB9FH410 (see AAW39155). The detection of
 CC such proteins and a CFH antigens can be used in screening or for the
 CC treatment of renal or urogenital cancer, e.g. bladder, cervical or
 CC prostate cancer. Agents that may modulate this antigen could be used in
 CC the manufacture of a medicament for the treatment of a tumour cell.
 XX
 SQ Sequence 240 AA;

Query Match 72.6%; Score 1362; DB 18; Length 240;
 Best Local Similarity 100.0%; Pred. No. 3.7e-97;
 Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 FTLTGNVFEYGVKAVYTCNEGYYQLLGEINRYRECDTGDGTNDIPICEVVKCLPVTAPENG 137
 |||||
 Db 1 fultgnvfeygvkavtycnegyqllegeinyrecdtgdwtndipicevvkclpvtapeng 60
 |||||
 QY 138 KIVSSAMEPDRYHFGQAVRFVNCNSGYKIEGDEEMHCSDDGFWSEKPKCVCISCKSPDV 197
 |||||
 Db 61 kivssamepdrayhfgqavrfvcnsgyklegdeemhcsddgfwsekpkcvcisckspdv 120
 |||||
 QY 198 INGSPIQKIIYKENERFOYKCNMGYEXSERGDVACTESGWRPLPSCSEKSCDNFYIPNG 257
 |||||
 Db 121 lngspisqkilykenerfyykcnmgyeysergdavctesgwrplpsceekscdnfyipng 180
 |||||
 QY 258 DYSPLRIKHTGDEITYQCRNGFYPATRGNTAKCTSTGWIAPAPRCTLPCDYPDIKHGGL 317
 |||||
 Db 181 dysplrikhtgdeityqcrngfypatrgntaktctstgwiapaprcctlkpcdypdikhagl 240

RESULT 2
 AAW39155
 ID AAW39155 standard; Protein: 216 AA.
 XX
 AC AAW39155;
 XX
 DT 27-APR-1998 (first entry)
 XX
 DE Clone PRB9FH410 CFH related protein fragment.
 XX
 KW Complement factor H; tumour associated antigen; renal cancer;
 KW urogenital cancer; medicament; modulator.
 XX
 OS Synthetic.
 XX
 PN WO9738136-A1.
 XX
 PD 16-OCT-1997.
 XX
 PF 09-APR-1997; 97WO-US05710.
 XX
 PR 06-MAR-1997; 97US-0812481.
 PR 09-APR-1996; 96US-0015083.
 PR 09-APR-1996; 96US-0630048.
 PR 06-MAR-1997; 97US-0038614.
 XX
 PA (BARD-) BARD DIAGNOSTIC SCI INC.
 XX
 XX Enfield DL, Hass GM, Kinders RJ;
 PI
 XX WPI: 1997-512742/47.
 DR N-PSDB; AAV02791.
 DR
 XX Treating or screening for cancer, e.g. renal or urogenital cancer -
 PT by modulating or detecting tumour associated human complement Factor

PT H related antigen, or nucleic acid encoding it
 XX
 PS Example 6B; Fig 6B; 104pp; English.
 XX
 CC This partial protein is found in clone PRB9FH410 and represents a
 CC complement factor H related protein with homology to a region of the
 CC human tumour-associated complement factor H (CFH). The detection of this
 CC protein and a CFH antigen can be used in screening or for the treatment
 CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.
 CC Agents that may modulate this antigen could be used in the manufacture of
 CC a medicament for the treatment of a tumour cell.
 XX
 SQ Sequence 216 AA;

Query Match 65.5%; Score 1229; DB 18; Length 216;
 Best Local Similarity 99.5%; Pred. No. 5.7e-87;
 Matches 215; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 95 TCNEGYYQLLGEINRYRECDTGDGTNDIPICEVVKCLPVTAPENGKIVSSAMEPDRYHFGQ 154
 |||||
 Db 1 tcnegyqllegeinyrecdtgdwtndipicevvkclpvtapengkivssamepdrayhfgq 60
 |||||
 QY 155 AVRFVNCNSGYKIEGDEEMHCSDDGFWSEKPKCVCISCKSPDVINGSPISQKIIYKENER 214
 |||||
 Db 61 avrfvcnsgyklegdeemhcsddgfwsekpkcvcisckspdvingspsqkilykener 120
 |||||
 QY 215 FOYKCNMGYEXSERGDVACTESGWRPLPSCSEKSCDNFYIPNGDYSPLRIKHTGDEITY 274
 |||||
 Db 121 foykcnmgyeysergdavctesgwrplpsceekscdnfyipngdysplrikhtgdeity 180
 |||||
 QY 275 QCRNGFYPATRGNTAKCTSTGWIAPAPRCTLPCDYP 310
 |||||
 Db 181 qcrngfypatrgntaktctstgwiapaprcctlkpcdyp 216

RESULT 3
 AAY55752
 ID AAY55752 standard; Protein: 496 AA.
 XX
 AC AAY55752;
 XX
 DT 22-FEB-2000 (first entry)
 XX
 DE Human CR1 protein LHR-A SCR fragment.
 XX
 KW C3B/C4B receptor; CR1 protein; cell-surface protein; erythrocyte; human;
 KW complement regulatory activity; complement pathway enzyme; tissue damage;
 KW reperfusion injury; Arthus reaction; myocardial infarct; inflammation;
 KW heart condition; autoimmune disorder; long homologous repeat; LHR; SCR;
 XX short consensus repeat.
 XX
 OS Homo sapiens.
 XX
 PN US5981481-A.
 XX
 PD 09-NOV-1999.
 XX
 PF 06-JUN-1995; 95US-0470652.
 XX
 PR 03-APR-1989; 89US-0332865.
 PR 06-DEC-1974; 74US-0350238.
 PR 24-FEB-1993; 93US-0026134.
 PR 01-APR-1988; 88US-0176532.
 XX
 PA (UYJO) UNIV JOHNS HOPKINS.
 XX
 PA (BGM) BRIGHAM & WOMENS HOSPITAL.
 PA (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.
 XX
 PI Concino MF, Wong WM, Makrides SC, Klickstein LB, Fearon DT, Ip SH;
 PI Marsh HC, Carson GR;
 XX
 XX WPI: 1999-633357/54.
 DR

XX A human C3B/C4B receptor (CRI) protein having antiinflammatory and
PT cardiant activity -
XX
XX
PS Disclosure; Fig 10; 87pp; English.
XX
CC The invention relates to a human C3B/C4B receptor (CRI) protein. The CRI
CC protein or fragment is expressed as a cell-surface protein on the surface
CC of a non-human cell and exhibits a complement regulatory activity of full
CC -length human CRI as expressed on erythrocytes. The CRI function in vivo
CC may be mediated through the inhibition of complement pathway enzymes. The
CC soluble CRI protein exhibits a complement regulatory activity, and this
CC may be used to prevent reperfusion injury, inhibit Arthus reaction, and
CC neutrophil mediated tissue damage, and reduce myocardial infarct size,
CC and inflammation. The CRI protein and its fragments can also be used in
CC the treatment of conditions which involve unwanted complement activity,
CC e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,
CC and autoimmune disorders. CRI proteins, analogues, derivatives, and anti
CC -CRI antibodies are used in assays, and diagnostics. The present sequence
CC represents the short consensus repeat (SCR) fragments of human CRI
CC protein long homologous repeat (LHR)-A sequence.
XX
SQ Sequence 496 AA;

Query Match 18.8%; Score 353; DB 20; Length 496;
Best Local Similarity 29.8%; Pred. No. 4.1e-19;
Matches 102; Conservative 54; Mismatches 126; Indels 60; Gaps 24;
QY 3 CNE---LPPRRNTEILTGSMDTYPEGTQAIYKCRPGYRSLGNVIMVCRKGE-WVALNP 58
DB 48 cnapewlpfarptnlt----defefpigtynyecrpgysgrpfslcclksnsvwtgkd 103
QY 59 LRKQCRPCGHPGDTFFGFTLFGNVFEGYKAVYTCNEGYQLLGEINYREC----DTD 114
DB 104 --rcrrkscrrnpdpvngmwhvkg--iqfsgqikysctkgyrligs--ssatciisgdtv 158
QY 115 GWNTDIPICEVVKC-LPVTAPENGKIVSSAMEPDRVHFGQAVFVCNSG-----YKIE 167
DB 159 iwdnetpdcdrpcglppti-tngdfistnre---nfhygsvvtyrcnpgsggrkvfelv 214
QY 168 GDEMHG---SDD--GFWSREKPKC-VEISCKSPDIVNGSPIS-OKIYKENERFOYKCNM 221
DB 215 gepsiyctsnddqvgiwsapqciipnkctppnvgilvsnrlsfnsvvefrccp 274
QY 222 GVEYSERGAVCTE-SGWRP-LPSCSEKSCNDPIPN-----GDYSPLRIKHRT 268
DB 275 gfvmkgprvrkqalnkweelpsc-srvccqpp--pdvIhaertqrkdndfsp----- 324
QY 269 GDEITVQCRNGFYPATRNTAKTSTG-WIP-APRCTLKPCD 308
DB 325 gqevfyscepg-ydlrgaasmrctpggdwspapaptcevkscd 365

RESULT 4
AAR13490
ID AAR13490 standard; Protein; 581 AA.
XX
AC AAR13490;

DT 30-OCT-1991 (first entry)
XX
DE Human C4 binding protein.

XX C4bp; monomer; complement protein; pJOD.C4bp.3; SCR;
KW short consensus repeat.
XX

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..32

FT /label= signal_peptide

FT Protein 33..581

FT Region /label= C4bp
FT 33..93 /label= SCR8
FT Region 94..155
FT /label= SCR7
FT 156..219
FT /label= SCR6
FT 220..279
FT /label= SCR5
FT Region 280..345
FT /label= SCR4
FT 346..406
FT /label= SCR3
FT Region 407..464
FT /label= SCR2
FT Region 465..523
FT /label= SCR1
FT Domain 524..581
FT /label= C4bp_core
FT /note= "responsible for multimer assembly"
FT Disulfide-bond 34..80
FT /note= "intradomain"
FT Disulfide-bond 65..92
FT /note= "intradomain"
FT
XX
PN W09111461-A.
XX
PD 08-AUG-1991.
XX
XX 28-JAN-1991; 91WO-US00567.
XX
PR 26-JAN-1990; 90US-0470888.
XX
PA (BIOG-) BIOGEN INC.
XX
PI Pasek MP, Winkler G, Liu TR;
XX
DR WPI; 1991-252613/34.
DR N-PSDB; AAR13242.
XX
PT New C4 binding protein fusion proteins and DNA encoding them -
PT comprise assemblies of C4bp monomers linked to functional moiety,
PT e.g. A2T, useful as delivery vehicles in diagnosis and therapy
XX
PS Example 1; Fig 1; 105pp; English.
XX
CC This sequence was deduced from human hepatocyte (Hep G2) cDNA
CC obtained following PCR amplification. The protein is a monomer
CC containing 8 SCRs. Each SCR forms a looped domain due to the
CC presence of two intradomain disulphide bonds (only the disulphide
CC bonds of SCR8 are labelled in the Features Table). Within each SCR,
CC the first cysteine residue bonds with the third and the second
CC cysteine residue bonds with the fourth. This secondary structure is
CC responsible for the conformational flexibility of the C4bp monomer.
CC The invention covers fusion proteins in which the monomer sequence,
CC or a specified subfragment of it, e.g. having 5, 4, 3 or 1 SCR(s)
CC is fused to the C-terminal of a protein such as a viral receptor,
CC cell ligand, a bacterial, viral or parasitic immunogen, enzyme,
CC cytokine, toxin, etc. See also AAR13243-51.
XX
SQ Sequence 581 AA;

Query Match 18.8%; Score 353; DB 12; Length 581;
Best Local Similarity 29.3%; Pred. No. 5e-19;
Matches 100; Conservative 50; Mismatches 141; Indels 50; Gaps 21;

QY 20 SDQTYPEGTQAIYKCRPGY-RSLGNVIMVCRK-GEWVALNPLKCRKPCGHPGDPFGT 77

DB 51 tetrfrktgttlktyclpgyvrshstqtlcnsdgeww-yntf--ciykrcrhpe----- 102

QY 78 FLTGTGNNV-----FEYGVKAVYTCNEGYQLLGEINYRECOTD-----GWNTDIPICEVVKC 128

Db 103 --lrngvvektldsfsgsfscseffllgsttstsr-cevqdrvgvshplpqceivkc 159
 Qy 129 LPVTAPENGKIVSSAMEPDRHYHFQAVRFVNCNSGYKIEGDEEHMCSD--GFWSKKEK 184
 Db 160 kppdirngrhsge----enfayagfsvtscdprfslghasictvenetigvwrp 215
 Qy 185 PKCWEISCKSPDVNGSPISO-KIIYKENERFQYKNCMGEYSESGDAV--C-TESGWRP 240
 Db 216 ptcekitcrkpdvshgenvsgfplnykdtivfkcdqkgfvl--rgssvihcdadsxw 273
 Qy 241 L-PSCEEKSCDN-PYIPNGDYS--PLRIK--HRTGDEITYOCNRYFPAT-RONTAKC- 291
 Db 274 sspacepncslndplphaswetyprptkedvyyvgtvirychpgykpdtdeptvvcq 333
 Qy 292 TSTGWIPAPRCTLKPCDPYDPIKHGGLYHENMRP-----YF 327
 Db 334 knlrwtpyqgcealccpeklnggeitqhrksrpanhcvyf 374

RESULT 5
 AAR28570
 ID AAR28570 standard; peptide; 543 AA.
 XX
 AC AAR28570;
 XX
 DT 19-MAR-1993 (first entry)
 XX
 DE CR1-4 (266-274 KLKTQTNASD) analogue.
 XX
 KW short consensus repeat; regulator of complement activation;
 KW C3b binding; C4b binding; human complement type 1 receptor.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 1..60
 FT /label= SCR-1
 FT Region 61..122
 FT /label= SCR-2
 FT Region 451..510
 FT /label= SCR-8
 FT Region 511..543
 FT /label= SCR-9
 FT /note= "TRUNCATED"
 FT Misc-difference 266..274
 FT /note= "ERTQRDN substituted with KLKTQTNASD
 FT from SCR-8-9 to increase C3b binding"
 XX
 PN EP512733-A.
 XX
 PD 11-NOV-1992.
 XX
 XX 28-APR-1992; 92EP-0303826.
 XX
 PR 03-MAY-1991; 91US-0695514.
 XX
 XX (UNIW) UNIV WASHINGTON.
 PA
 XX Atkinson JP, Hourcade D, Krych M;
 PI WPI; 1992-375009/46.
 XX
 DR Complement activity regulator protein analogues - useful for
 XX treating auto-immune diseases, to suppress transplant rejection,
 PT for diagnosis etc.
 PT
 XX
 XX Example 8; Page 18 and R11810; 23pp; English.
 PS
 CC The cDNA clone designated CR1-4 was described in J.Exp.Med.(1988)
 CC 168:1255-1270. It encodes the first 8 and a half amino terminal
 CC SCR8 of CR1. The invention concerns analogues of "regulator of
 CC complement activation" proteins or truncated, hybrid or recombinant
 CC forms of them. CR1-4 is a preferred truncated form and a number of

CC specified substitution variants of it are disclosed in which certain
 CC positions in SCR-5-6 are substituted by amino acids from
 CC the corresponding positions in SCR8 which are involved in C3b- and
 CC C4b-binding. The substitution variant given here has increased C3b-
 CC binding. The specification does not contain the CR1-4 sequence;
 CC the sequence given here was constructed from the full-length CR1
 CC amino acid sequence having GENESEQ accession number AAR11810 and
 CC descriptions in the disclosure.
 XX
 SQ Sequence 543 AA;
 Query Match 18.8%; Score 352.5; DB 13; Length 543;
 Best Local Similarity 29.6%; Pred. No. 5e-19;
 Matches 101; Conservative 53; Mismatches 128; Indels 59; Gaps 23;
 Qy 3 CNE---LPPRNRTEILTGSMSDQTYPEGTQAIYKCRGYSGLGNVIMVCRKGEWVALNPL 59
 Db 2 cnapewlpfarptnlt----defefpigtynycrpgysgrpfisilcknsvwtgakt- 56
 Qy 60 KCKOKRPGCHPGDTPFGFTLTGCVNFVEYGVKAVYTCNEGYSQLLGEINRYREC---D 115
 Db 57 -rcrrksrnpdpvngmvhvikg--lqfsgqlkysckkyrligs-ssatclisgdcvi 112
 Qy 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDRHYHFQAVRFVNCNSG-----YK 168
 Db 113 wdnetpicdrpcglppti-tngdfistnre---nfhygsvvtyrcnpgsggrkvfeivg 168
 Qy 169 DEEMHC--SDD--GFWSKEKPKC-VEISCKSPDVINGSPIS-QKIIYKENERFQYKCMG 222
 Db 169 epslyctsndqvgiwsppapqclipnktppnvengilvsnrslfslnevefrccpv 228
 Qy 223 YEYSERGDAVCTE-SGWRP-LPSCEEKSCDNPIYP-----NGDYSPLRIKHRTG 269
 Db 229 fvmkgprvkcqalnkweplpsc-srvccpp--pdvlhaklkqtasdtsp-----g 278
 Qy 270 DEITYOCNRYFPATRGNTAKTSTG-WIP-APRCTLKPCD 308
 Db 279 qevfyscepg-ydlrgaasmrctpqgdwspaaptcevkscd 318

RESULT 6
 AAW73147
 ID AAW73147 standard; protein; 778 AA.
 XX
 AC AAW73147;
 XX
 DT 29-JAN-1999 (first entry)
 XX
 DE Amino acid sequence of the soluble complement receptor 1 (sCR1).
 XX
 KW Human; soluble complement receptor 1; sCR1; T-cell; B-cell;
 KW mediated immune response; inhibition; tissue rejection; gene therapy;
 KW dystrophin; inflammatory response; interferon-gamma secretory response;
 KW autoimmune response; neurological response; Alzheimer's disease;
 KW Parkinson's disease; multiple sclerosis; systemic lupus erythematosus;
 KW rheumatoid arthritis; myasthenia gravis; epidermis bullosa;
 KW Hashimoto's disease.
 XX
 OS Homo sapiens.
 XX
 PN W09845430-A1.
 XX
 PD 15-OCT-1998.
 XX
 XX 06-APR-1998; 98WO-GB01012.
 XX
 PR 05-APR-1997; 97GB-0006950.
 XX
 XX (ANNE/) ANNENKOV A.
 PA (CHER/) CHERNAJOVSKY Y.
 XX
 PI Annenkov A, Chernajovsky Y;

XX	WPI; 1998-568350/48.
XX	
DR	
XX	
PT	Treating of soluble human complement receptor 1 - useful for
PT	fracturing T-cell or B-cell mediated immune responses e.g.
PT	inflammatory responses such as rheumatoid arthritis
XX	
PS	Disclosure; Fig 1; 54pp; English.
XX	
CC	This is an amino acid sequence of the human soluble complement
CC	receptor 1 (sCR1), useful in the treatment of T-cell or B-cell
CC	mediated immune responses. It is used to inhibit a T-cell or
CC	B-cell-mediated immune response to prevent immune response-mediated
CC	tissue rejection and destruction or clearance or inactivation of an
CC	expressed protein especially from cells that have been treated by gene
CC	therapy to express the protein, e.g. dystrophin. The protein can also
CC	be used to inhibit a T-cell or B-cell-mediated inflammatory response,
CC	an interferon-gamma secretory response, autoimmune response or
CC	neurological response, e.g. Alzheimer's or Parkinson's disease or
CC	multiple sclerosis. Also the protein can be used to treat systemic
CC	lupus erythematosus, rheumatoid arthritis, myasthenia gravis,
CC	epidermis bullosa or Hashimoto's disease.
XX	
SO	Sequence 778 AA:

```

Query Match      18.7%   Score 350.5;   DB 19; Length 778;
Best Local Similarity 29.3%; Pred. No. 1.1e-18;
Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;
```

RESULT	7	
AAW45899		
ID	AAW45899 standard; peptide; 1930 AA.	
XX		
XX		
XX	AAW45899;	
XX		
DT	30-JUN-1998 (first entry)	
XX		
XX		
DE	Human complement receptor 1 (residues 1-1929).	
XX		
XX		
KW	Membrane binding element; thrombotic disease; soluble protein;	
KW	complement-related disease; integral membrane protein; inflammation;	
KW	short consensus repeat; SCR 1-3; CR1; complement receptor type 1.	
XX		
OS	Homo sapiens.	

xx	Key	Location/Qualifiers
FH		1930
FT	Cross-links	

FT	/note= "Disulphide linked to Cys in peptide given
FT	in AAW4589"
XX	
XX	WO9802454-A2.
XX	
PD	22-JAN-1998.
XX	
XX	08-JUL-1997; 97WO-EP03715.
XX	
XX	15-JUL-1996; 96GB-0014871.
XX	
XX	(ADPR-) ADPROTECH PLC.
XX	
XX	Dodd I, Mossakowska DEI, Smith RAG;
PI	
XX	WPI; 1998-110524/10.
DR	
XX	
PT	Derivatives of soluble poly:peptide(s) bonded to low affinity
PT	membrane binding groups - useful for treating complement-related and
PT	thrombotic diseases, providing improved localisation at cellular
PT	membranes
XX	
XX	Claim 22; Pages 60-61; 75pp; English.
PS	
XX	
CC	This sequence represents human complement receptor 1 (CR1, CD 35)
CC	N-terminal fragment. The invention relates to a soluble derivative (A)
CC	of a soluble polypeptide (I), which comprises at least 2 heterologous
CC	membrane-binding elements (MBE) of low membrane affinity covalently
CC	associated with (I). MBE interact, independently and with thermodynamic
CC	additivity, with components of cellular or artificial membranes exposed
CC	to extracellular fluids. (A) are used to treat disorders treatable with
CC	(I) itself, specifically inflammation or any other complement-related
CC	disorder (e.g. neurological disease, graft rejection, myocardial
CC	infarction, sepsis, rheumatoid arthritis and many others; including
CC	application to indwelling devices) and thrombolytic disease, but also to
CC	treat allergy, induce weight loss, to treat ischaemia or asthma and as
CC	immuno-modulators for treating multiple sclerosis. (A) are administered
CC	orally, topically, by injection or inhalation at 0.01-10 (preferably
CC	0.1-10) mg/kg/day.
XX	
XX	Sequence 1930 AA:
SQ	

```

Query Match      18.7%   Score 350.5;   DB 19;   Length 1930;
Best Local Similarity 29.3%;   Pred. No. 3.3e-18;
Matches 100;   Conservative 55;   Mismatches 127;   Indels 59;   Gaps 23;

QY      3 CNE---LPPRNTETILGCSNSDQTYPECTQAIYKCRPGYSLGNVIMYCRKGWVALNPL 59
       || | | | | : | | | | | | | | | | : | | | | | | | | | |
Db      2 cnapewlpfarptnlt-----defeflgtlylnyecrpgysgrfsilclknsvvtgaxd- 56
       || | | | | : | | | | | | | | | | : | | | | | | | | | |

QY     60 RKCKRPCHGPDPNPFGTFTLTGNGVFYEYGVKAVVTGNEGYYOLLGEINVRBC----DFDG 115
       || | | | | : | | | | | | | | | | : | | | | | | | | | |
Db     57 -rcrrkcrnpdpvgmvmviki--lqfsgqlkysectgyrligs-ssatcilsgdvlvi 112
       || | | | | : | | | | | | | | | | : | | | | | | | | | |

QY    116 WTNDIPICEVKKC-LPVTAPENGKIVSAMBPDRHYHFGQAVRFVCNSG-----YKIEG 168
       || | | | | : | | | | | | | | | | : | | | | | | | | | |
Db    113 wdnetpicdrpcglpbpti-tngafistnr--nfnygsvvtyrcnpgsggrkvfelvg 168
       || | | | | : | | | | | | | | | | : | | | | | | | | | |

QY    169 DEEMHC--SDD-GFWSKEPKC-VEISCKSPDVINGSPIS-QKIIVKENERFOYKCNMG 222
       || | | | | : | | | | | | | | | | : | | | | | | | | | |
Db    169 epslyctsnddqgiwggppaqcplpnktcpnvengllvsdnrsifslnevfvefrccpq 228
       || | | | | : | | | | | | | | | | : | | | | | | | | | |

QY    223 YEYSERGDVACTE-SGWRP-LPSCSEERSCONPYLPN-----GDYSPRIKURTG 269
       || | | | | : | | | | | | | | | | : | | | | | | | | | |
Db    229 fvmkgprvrvcqalnkweelpsc-srvccpp--pdvlhaertqrkdnfsp-----g 278
       || | | | | : | | | | | | | | | | : | | | | | | | | | |

QY    270 DETIYQCNRGFYPATRGNTAKCTSTG-WIP-APRCTLKPCD 308
       || | | | | : | | | | | | | | | | : | | | | | | | | | |
Db    279 qevfysecpq-ydlrqasmrcttqqgdwsaaaptcevkscd 318
       || | | | | : | | | | | | | | | | : | | | | | | | | | |

```

RESULT 8

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

Sequence 2039 AA;

Query Match	18.7%	Score	350.5	DB	22	Length	2039
Best Local Similarity	29.3%	Pred. No.	3.6e-18				
Matches	100	Conservative	55	Mismatches	12	Indels	59
						Gaps	23

[illegible]

RESULT 10

ABB11782
ID ABB11782 standard; peptide: 2044 AA.

AC ABB11782;

DT 11-JAN-2002 (first entry)

Human CR1 protein homologue, SEQ ID NO:2152.

KW	Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW	haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW	inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW	proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW	myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW	chronic inflammatory condition; proliferative retinopathy;
KW	atherosclerosis; coronary heart disease; arterial ischaemia;
KW	bone disorder; osteoporosis; vascular growth disorder;
KW	tissue regeneration; wound healing; infection; immune disorder;
KW	cell culture; drug screening; gene therapy; antiinflammatory;
KW	antiasthmatic; antiarthritis; haemostatic; antiarteriosclerotic;
KW	cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW	antifungal; vulnery; antileucic.
KW	

OS Homo sapiens.

PN WO200157188-A2.

PD 09-AUG-2001.

PF 05-FEB-2001; 2001WO-US03800.

PR 03-FEB-2000; 2000US-0496914.

XX

XX
XX

PI Tang YT, Liu C, Drmanac RT;

[illegible]

Query Match 18.7%; Score 350.5; DB 22; Length 2044;
Best Local Similarity 29.3%; Pred. No. 3.6e-18;
Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;
3 CNE---LPRRNTEILTGSWSDOTPEGTQATYKCRPGYRSIGNVIMCRKGEWALNPL 59

Qy 3 CNE---LPPRNTILTGSWSDQTYPGTQTATYKCRPGYSLSGNVIMVCRGGEWALNPL 59
|| || | : : ||| : ||| : :
Db 48 cnapewlpfarptnlt----defefglqtlnyecrpgysgrpfslclcnsvwtakd- 102

Qy	60	RKQKRPCGHPGDPPFGTFTLTGGNNFEYGVKAVVTCNEGYQLLGEIN	REC---DTDG	115
		: : : : : : : : : : : : : : : :		
Dd	103	-rcrrkscrnppdpvngmvhlvg--lqfsgqlkysectgyrligs-ssatclisgdtvi		158

Qy 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDREYHFGQAVRRCVNSG-----YKIEG 168

Db 159 wdnetspdrpcglppti-tngdfistnre---nfhygsvvtvrcnpgsgggrkvfelvg 214

	: :: : : : : : : : :	
Qy	169 DEEMHC--SDD--CFWSKERPKC-VEISCKSPDVINGSPIS-QXILYKERNERFQVKNMG	222
Dd	215 epslyctsndddqvgilwsgpapqcifpnkctppnvengllvdsndrslfsinevvfrfcpg	274

QY 223 YEYSERGDAVCTE-SGWRP-LPSCBEKSCDNPYIPN-----GDYSPLEIKHRTG 269

WPI; 2001-457740/49.
N-PSDB; ABA09026.

Human proteins and DNA encoding sequences useful for preventing, treating or ameliorating a medical condition in a mammalian subject e.g. arthritis and cancer -

Claim 20; Page 245-246; 1963pp; English.

Sequences ABBL0981-ABBL2330 represent 1350 novel human polypeptides, and sequences ABA08225-ABA09574 represent nucleic acids encoding them. The invention also relates to vectors and recombinant host cells comprising a nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of detecting the nucleotides or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoiesis regulatory activity; tissue growth activity;

immunomodulatory activity; activin- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g. by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a novel human polypeptide of the invention.

Sequence 2044 AA:

Query Match	18.7%;	Score 350.5;	DB 22;	Length 2044;
Best Local Similarity	29.3%;	Pred. No. 3.6e-18;		

Qy 3 CNE---LPPRRNTEILTGSWSDQTYPEGTAIYKCRPGYSLSGNIWMVCRKGEWALNPL 59
|| || | : : ||| : ||| : :
Db 48 cnapewlpfartnlt----defeflqtlvncvrpgvsgrpfslclcnsvwtakd- 102

Qy	60	RKQKRPCGHPGDPPFGTFTLTGGNNFEYGVKAVVTCNEGYQLLGEIN	REC---DTDG	115
		: : : : : : : : : : : : : : :		
Dd	103	-rcrrkscrnppdpvngmvhlvg--lqfsgqlkysectgyrligs-ssatclisgdtvi		158

Qy 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDREYHFGQAVRRCVNSG-----YKIEG 168

Db 159 wdnetspdrpcglppti-tngdfistnre---nfhygsvvtvrcnpgsggrkvfelvg 214

Qy	169	DEEMHC--SDD--CFWSKERPKC-VEISCKSPDIVNGSPIS-QKILYKENERFOYKCNMG	222
Db	215	eps1yctsnddqvglwsgpappqclfnkctppnvengllvdsdnrlsflsinevvefrccpqg	274

QY 223 YEYSERGDAVCTE-SGWRP-LPSCBEKSCDNPYIPN-----GDYSPLEIKHRTG 269

XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 XX Example 2; SEQ ID NO 5941; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AA138642-AA142213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 2044 AA;

Query Match 18.7%; Score 350.5; DB 22; Length 2044;
 Best Local Similarity 29.3%; Pred. No. 3.6e-18;
 Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;
 QY 3 CNE---LPRRRTELTGWSQDQTYEGTQAIYKCRPGYRSLGNVIMVCRKGEWALNPL 59
 DB 48 napewlpfartnlt----defefpigtlynyecrpgysgrpfslclcknsvwtgkd- 102
 QY 60 RKCKRQPCGHFGDTPFGTFTLTGGNVFYGKAVYTCNEGQYLLGEINRYREC----DTDG 115
 DB 103 -rcrrkscrnppdpvngmvhvikg--iqfgsqikysctckgyrligs-ssatclisgdtvi 158
 QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDRYHFGQAVRFVCSNG-----YKIEG 168
 DB 159 wdnetspdrpcglppti-tngdfistnre---nfhygsvvtyrcnpgsggrkvfelvg 214
 QY 169 DEEMHC--SDD--GFWSKEPKC-VEISCKSPDVINGSPIS-OKIYKENERFOYKCNMG 222
 DB 215 epslyctsnddqvglwsgpapqciipnktppnvengllvsdnrlslfslneveifrcpqp 274
 QY 223 YEYSERDAVCTE-SGWRP-LPSCEEKSCDNPYIPN-----GDYSPLRIKHRTG 269
 DB 275 fvmkgprrvkcqalnkwepeipsc-srvccqp--pdvlhaertqrkdnfsp-----g 324
 QY 270 DEITYQCRNGFYPATRGNTAKTSTG-WIP-APRCTLKPCD 308
 DB 325 qevfyscepg-ydlrgaasmrctpggdwspaaptcevkscd 364

RESULT 13
 AAP92219
 ID AAP92219 standard; protein; 2317 AA.
 XX
 AC AAP92219;
 XX

22-FEB-1990 (first entry)
 DT
 XX
 DE CRI protein.
 XX
 KW Complement; cofactor.
 XX
 OS Homo sapiens (human).
 XX
 FH Key Location/Qualifiers
 FT Peptide 10..50
 FT /label= signal_peptide
 XX

PN W08909220-A.
 XX
 PD 05-OCT-1989.
 XX
 PF 31-MAR-1989; 89WO-US01358.
 XX
 PR 01-APR-1988; 88US-0176532.
 XX
 PA (TCEL) T CELL SCIENCES INC.
 PA (UJO) THE JOHNS HOPKINS UNIVERSITY.
 PA (BRIG) THE BRIGHAM AND WOMEN'S HOSPITAL.
 XX
 PI Fearon DT, Klickstein LB, Wong W, Carson G, Concino MF, Makrides SC;
 XX WPI; 1989-309498/42.
 DR N-PSDB; AAN91477.
 XX
 PT New nucleic acid sequences encoding new CRI protein - and its fragment,
 PT for diagnosis and control of complement-related immune defects,
 PT inflammation, myocardial infarct, etc
 XX
 PS Claim 1; Fig 1; 191pp; English.
 XX
 CC This is full-length CRI protein, and shortened forms are new, lacking
 CC the transmembrane region. The proteins and fragments bind C3b and/or
 CC C4b, have cofactor activity and inhibit C3 and C5 convertase activity.
 CC In the sequence, x=untranslated region. This has 7 short consensus
 CC repeats (CR) of 4 long direct homologous repeats (LHR) (see fig. 10 in
 CC patent). Deletion mutants, eg LHR-A for binding of C4b, and LHR-C
 CC for C3b binding, be made, and LHR-B and -C can detect CR2 sequences.
 CC They are useful in diagnosing and treating immune disorders, and prevent
 CC perfusion injury.
 XX
 SQ Sequence 2317 AA;

Query Match 18.7%; Score 350.5; DB 10; Length 2317;
 Best Local Similarity 29.3%; Pred. No. 4.1e-18;
 Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;
 QY 3 CNE---LPRRRTELTGWSQDQTYEGTQAIYKCRPGYRSLGNVIMVCRKGEWALNPL 59
 DB 52 napewlpfartnlt----defefpigtlynyecrpgysgrpfslclcknsvwtgkd- 106
 QY 60 RKCKRQPCGHFGDTPFGTFTLTGGNVFYGKAVYTCNEGQYLLGEINRYREC----DTDG 115
 DB 107 -rcrrkscrnppdpvngmvhvikg--iqfgsqikysctckgyrligs-ssatclisgdtvi 162
 QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDRYHFGQAVRFVCSNG-----YKIEG 168
 DB 163 wdnetspdrpcglppti-tngdfistnre---nfhygsvvtyrcnpgsggrkvfelvg 218
 QY 169 DEEMHC--SDD--GFWSKEPKC-VEISCKSPDVINGSPIS-OKIYKENERFOYKCNMG 222
 DB 219 epslyctsnddqvglwsgpapqciipnktppnvengllvsdnrlslfslneveifrcpqp 278
 QY 223 YEYSERDAVCTE-SGWRP-LPSCEEKSCDNPYIPN-----GDYSPLRIKHRTG 269
 DB 279 fvmkgprrvkcqalnkwepeipsc-srvccqp--pdvlhaertqrkdnfsp-----g 328
 QY 270 DEITYQCRNGFYPATRGNTAKTSTG-WIP-APRCTLKPCD 308
 DB 329 qevfyscepg-ydlrgaasmrctpggdwspaaptcevkscd 368

RESULT 14
 AAB53125
 ID AAB53125 standard; Protein; 645 AA.
 XX
 AC AAB53125;
 XX
 DT 28-FEB-2001 (first entry)
 XX

DE Macaca mulatta rhadinovirus 17577 RRV ORF4 protein SEQ ID NO:7.
 XX
 KW genome: Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;
 KW IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;
 KW cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;
 KW lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
 KW splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia.
 XX
 OS Macaca mulatta rhadinovirus 17577.
 XX
 PN WO200028040-A2.
 XX
 PD 18-MAY-2000.
 XX
 PF 05-NOV-1999; 99WO-US26260.
 XX
 PF 06-NOV-1998; 98US-0107507.
 XX
 PR 20-NOV-1998; 98US-0109409.
 XX
 PA (UYOR-) UNIV OREGON HEALTH SCI.
 XX
 PI Wong SW, Athelm MK, Searles RP;
 XX
 DR WPI; 2000-376552/32.
 XX
 PT New rhesus rhadino virus for producing non-human primate model useful
 PT for testing potential treatments and efficacy of the candidate vaccine
 PT for conditions associated with RRV infection -
 XX
 PS Claim 5; Page 122-123; 141pp; English.
 XX
 CC The present invention describes a novel rhesus macaque rhadinovirus
 CC called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the
 CC RRV genome sequence, and AAB53123 to AAB53204 represent the proteins
 CC encoded by the genome sequence. The present invention also specifically
 CC claims the individual open reading frame (ORF) nucleotide sequences from
 CC the genome which encode the individual proteins, but these sequences are
 CC not given. A non-human animal infected with RRV can be used for testing
 CC the efficacy of drug in the treatment of condition associated with
 CC infection with RRV such as Kaposi's sarcoma, lymphoproliferative
 CC disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,
 CC hypergammaglobulinemia or autoimmune haemolytic anaemia, by
 CC administering the drug to a immuno-compromised non-human primate
 CC preferably Rhesus macaque monkey obtained by as a result of infection
 CC by Simian Immunodeficiency Virus (SIV). RRV is useful for producing
 CC non-human primate model for testing potential treatments for conditions
 CC associated with RRV infection. It is also useful for testing the
 CC efficacy of the candidate vaccine against RRV infection or conditions
 CC associated with its infection by administering the vaccine to the
 CC subject capable of infection with RRV, inoculating the subject with RRV
 CC and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205
 CC to AAB53213 represent sequence used in the exemplification of the
 CC present invention.
 XX
 SQ Sequence 645 AA;

Query Match 18.6%; Score 349.5; DB 21; Length 645;
 Best Local Similarity 25.8%; Pred. No. 1.1e-18;
 Matches 99; Conservative 45; Mismatches 164; Indels 75; Gaps 14;

Qy 1 EDCNELPPRRNTEILTGSWSQDTPEGTQAIYKCRPGYRSLGNVIMV--CRKGWALNP 58
 Db 21 encK--pphftfeyrvksntekdlvsgetaelicrpgyvtntkltiteclngtw--stp 76
 Qy 59 LRKCKRCPGHPGTPTFTLTGG-VNFEYGVRAVTCNTEGYOLLGEINPREC---DTD 114
 Db 77 nfpcdkrctpadllngavhthghdhaldfgmsisynecngdlgs-nvrflclgqte 135
 Qy 115 --GWTNDIPICEVYKCLPVTAPENGKIVSSAMEPDRE-YHFGQAVRFVNSGYKIEGDEE 171
 Db 136 nvnwdsnepvcqlcklppavehgdyl-----pnqdvnyngdaiftkcsisylcivgst 190

Qy 172 MHCSDGFWSEKPKCEVLSCKSPDVING-SPISQKIIYKENERFOYKCNMGVEYSERGD 230
 Db 191 ivctsnkkswnsfptclmivcespqldngyldlglsrryhnghqstlvcscdgynlvgpet 250
 Qy 231 AVCTESGW-RPLPSC-----EKSQDNPIYPNGDYSPLRI----- 264
 Db 251 lctctnttwvpplpkcvlvcnnpstmpetpmpetpdyqklnlstaktattpnafvtv 310
 Qy 265 -----KURTGDEITYQCRNGFYPATRGNTAKCTSTGWI 297
 Db 311 vspekddvtcvkphferfmvkaendekysvgasvelicrpgftkmgstvsveclngtw 370
 Qy 298 PAP--RCTLKPCDYPDIKHGGLY 318
 Db 371 tapnakchrkkcptpqellingey 393

RESULT 15
 AAR28557
 ID AAR28557 standard; peptide; 543 AA.
 XX
 AC AAR28557;
 XX
 DT 19-MAR-1993 (first entry)
 XX
 DE CR1-4 (99H, 103E) analogue.
 XX
 KW short consensus repeat; regulator of complement activation;
 KW C3b binding; C4b binding; human complement type 1 receptor.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 1..60
 FT /label= SCR-1
 FT Region 61..122
 FT /label= SCR-2
 FT Region 451..510
 FT /label= SCR-8
 FT Region 511..543
 FT /label= SCR-9
 FT /note= "TRUNCATED"
 FT Misc-difference 99
 FT /note= "Ser substituted by His (SCR-9)"
 FT Misc-difference 103
 FT /note= "Thr substituted by Glu (SCR-9)"
 XX
 PN EP512733-A.
 XX
 PD 11-NOV-1992.
 XX
 PF 28-APR-1992; 92EP-0303826.
 XX
 PR 03-MAY-1991; 91US-0695514.
 XX
 PA (UNIW) UNIV WASHINGTON.
 XX
 PI Atkinson JP, Hourcade D, Krych M;
 XX
 DR WPI; 1992-375009/46.
 XX
 PT Complement activity regulator protein analogues - useful for
 PT treating auto-immune diseases, to suppress transplant rejection,
 PT for diagnosis etc.
 XX
 PS Claim 11; Fig 2 and R11810; 23pp; English.
 XX
 CC The cDNA clone designated CR1-4 was described in J.Exp.Med.(1988)
 CC 168:1255-1270. It encodes the first 8 and a half amino terminal
 CC SCRs of CR1. The invention concerns analogues of "regulator of
 CC complement activation" proteins or truncated, hybrid or recombinant
 CC forms of them. CR1-4 is a preferred truncated form and a number of

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: August 29, 2002, 15:02:17 : Search time 39.15 Seconds
(without alignments)
1050.479 Million cell updates/sec

Title: US-09-316-163-14

Perfect score: 2454

Sequence: 1 EDCKGPPPRENSILSGWS.....DTYYCTENGWSPPPKCVRIK 428

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2055	83.7	1234	1	NBM5H	complement factor
2	1714	69.8	1231	1	NBH0H	complement factor
3	1709	69.6	449	1	NBH0H	complement factor
4	1242	50.6	669	2	S6551	factor H - bovine
5	813	33.1	452	2	A35068	complement factor
6	811	33.0	303	2	H35068	apolipoprotein H-r
7	569.5	23.2	1053	2	S46199	probable complement
8	566	23.1	808	2	D35069	complement factor
9	529.5	21.6	597	1	S53711	C4BP alpha chain p
10	481.5	19.6	597	1	NBHUC4	C4b-binding protei
11	474.5	19.3	610	1	I46001	C4b-binding protei
12	466.5	19.0	331	2	A45222	complement factor
13	453	18.5	676	2	A45900	complement C3b rec
14	449	18.3	558	2	S79933	C4BP protein alpha
15	447	18.2	2014	2	I36936	complement recepto
16	437	17.8	2489	2	I73012	complement C3b/C4b
17	409	16.7	497	2	JC2054	complement regulat
18	400.5	16.3	482	2	A34924	complement C3b/C4b
19	396	16.1	579	2	A56740	Sperm-egg recognit
20	384.5	15.7	469	1	NBM5C4	C4b-binding protei
21	375	15.3	560	2	I16833	hypothetical prote
22	373	15.2	1025	1	A43526	complement C3d/Eps
23	363	14.8	1091	1	PL0009	complement C3d/Eps
24	336	13.7	661	1	KFHU13	coagulation factor
25	332	13.5	668	2	A46013	coagulation factor
26	319	13.0	343	2	G35070	apolipoprotein H-r
27	318	13.0	440	2	A43519	complement recepto
28	315	12.8	330	2	I56100	complement factor
29	314.5	12.8	263	1	WMVZSP	apolipoprotein H h

ALIGNMENTS

RESULT 1

NBM5H

complement factor H precursor - mouse

N:Alternate names: protein beta-1-H

C:Species: Mus musculus (house mouse)

C>Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 22-Jun-1999

C:Accession: A26154; I49711; I49728

R:Kristensen, T.; Tack, B.F.

Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986

A:Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in lengt

A:Reference number: A26154; MUID:86233353

A:Accession: A26154

A:Molecule type: mRNA

A:Residues: 1-1234 <KRI>

A:Cross-references: GB:M2660; NID:gl93724; PIDN:AAA37759.1; PID:g387181

R:Natsuume-Sakai, S.; Nonaka, M.; Nonaka, M.; Harada, Y.

J. Immunol. 144, 358-362, 1990

A:Title: Demonstration of an unusual allelic variation of mouse factor H by the compl

A:Reference number: I49711; MUID:90111033

A:Accession: I49711

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-18 <RES>

A:Cross-references: GB:M31979; NID:gl93726; PIDN:AAA37762.1; PID:g193729

R:Munoz-Canoves, P.; Tack, B.F.; Vlk, D.P.

Biochemistry 28, 9891-9897, 1989

A:Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma

A:Reference number: I49728; MUID:90148935

A:Accession: I49728

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-19 <RE2>

A:Cross-references: GB:J02891; NID:gl93805; PIDN:AAA37795.1; PID:g553926

C:Comment: Two codominant alleles of factor H are present in mice.

C:Comment: Factor H functions as a cofactor in the inactivation of C3b by serine prot

(C5 convertase) in the alternative complement pathway.

C:Genetics:

A:Map position: 1

C:Superfamily: complement factor H; complement factor H repeat homology

C:Keywords: complement alternate pathway; duplication; glycoprotein; plasma

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-1234/Product: complement factor H #status predicted <MPT>

F:21-80/Domain: complement factor H repeat homology <FH01>

F:85-141/Domain: complement factor H repeat homology <FH02>

F:146-205/Domain: complement factor H repeat homology <FH03>

F:210-262/Domain: complement factor H repeat homology <FH04>

F:246-248/Region: cell attachment (R-G-D) motif

F:267-320/Domain: complement factor H repeat homology <FH05>

F:325-385/Domain: complement factor H repeat homology <FH06>

F:389-442/Domain: complement factor H repeat homology <FH07>

F:448-505/Domain: complement factor H repeat homology <FH08>

F:509-564/Domain: complement factor H repeat homology <FH09>


```
F:569-622/Domain: complement factor H repeat homology <FH10>
F:629-683/Domain: complement factor H repeat homology <FH11>
F:690-743/Domain: complement factor H repeat homology <FH12>
F:752-802/Domain: complement factor H repeat homology <FH13>
F:808-861/Domain: complement factor H repeat homology <FH14>
F:867-931/Domain: complement factor H repeat homology <FH15>
F:936-989/Domain: complement factor H repeat homology <FH16>
F:994-1048/Domain: complement factor H repeat homology <FH17>
F:1053-1107/Domain: complement factor H repeat homology <FH18>
F:1114-1168/Domain: complement factor H repeat homology <FH19>
F:1172-1233/Domain: complement factor H repeat homology <FH20>
F:21-66-52-80-85-129-114-141-146-192-178-205-210-251-237-262-267-309-294-320-325-374-357
O-802,808-850,836-861,867-920,906-931,936-976,964-989,994-1037,1023-1048,1053-1096,1082-
F:676,721,773,801,1030,1061,1225/Binding site: carbohydrate (Asn) (covalent) #status pre

Query Match      83.7%: Score 2055; DB 1; Length 1234;
Best Local Similarity 82.9%: Pred No. 1.4e-135;
Matches 355; Conservative 28; Mismatches 45; Indels 0; Gaps 0;

QY 1  EDCKGPPPRENSEILLGSNSQYSEGTOATYKCRPGYRTLTGTVKCKNGEWPSPNPSR 60
DB 19  EDCKGPPPRENSEILLGSNSQYSEGTOATYKCRPGYRTLTGTVKCKNGKWAASNPSPR 78
QY 61  ICRKPCGHPGDPFGSPRLAVGSEFFERCAKVVYTCDEGYQLLGEIDYRECDAGWTNDI 120
DB 79  ICRKPCGHPGDPFGSPRLAVGSEFFERCAKVVYTCDEGYQLLGEIDYRECDAGWTNDI 138
QY 121  PICEVVKCLPVTLENGRIVTSCAASPDOBYEFGQVYVRFECNKGFKTEGOKEMHCSENGLW 180
DB 139  PICEVVKCLPVTLENGRIVTSCAASPDOBYEFGQVYVRFECNKGFKTEGOKEMHCSENGLW 198
QY 181  SNEKPCQVEISCLPPRVENGDIYLKPVYKNERFYKCKQGVYKRGSDAVCTGSGWNP 240
DB 199  SNEKPCQVEISCLPPRVENGDIYLKPVYKNERFYKCKQGVYKRGSDAVCTGSGWNP 258
QY 241  QPSCSEMTCLTPYINGIYTPRIKRIHDEIRYCKNGFYFATSPSKYCKITGWIPAP 300
DB 259  QPSCSEMTCLTPYINGIYTPRIKRIHDEIRYCKNGFYFATSPSKYCKITGWIPAP 318
QY 301  RGLSLPCDFPQPKHGLRYEESRRPYFPVPIGKEYSYXCDNGFTTPSQSYWDYLRCTVNG 360
DB 319  RGLSLPCDFPQPKHGLRYEESRRPYFPVPIGKEYSYXCDNGFTTPSQSYWDYLRCTVNG 378
QY 361  WEPEVPCVLRQCLFHYVEGESSYWRRIEYGSAKQVCHSGYSLNPGQDTYCTENGWSP 420
DB 379  WEPEVPCVLRQCLFHYVEGESSYWRRIEYGSAKQVCHSGYSLNPGQDTYCTENGWSP 438
QY 421  PPKCVRIK 428
DB 439  PPKCIRIK 446

RESULT 2
NBHUH
complement factor H precursor, long splice form [validated] - human
C:Species: Homo sapiens (man)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 08-Dec-2000
C:Accession: S00254; A60238; A54726; A61565; A26505; I72654; S66298
R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.
Biochem. J. 249, 593-602, 1988
A:Title: The complete amino acid sequence of human complement factor H.
A:Reference number: S00254; MUID:88134059
A:Accession: S00254
A:Molecule type: mRNA
A:Residues: 1-1231 <RIP>
A:Cross-references: EMBL:Y00716; NID:g31964; PIDN:CAA68704.1; PID:g31965
A:Note: 402-Tyr was also found
A:Note: parts of this sequence, including the amino and carboxyl ends of the mature protein, are identical to the sequence of factor H from sheep.
R:Estallier, C.; Schwaeble, W.; Dierich, M.; Weiss, E.H.
Eur. J. Immunol. 21, 799-802, 1991
A:Title: Human complement factor H: two factor H proteins are derived from alternatively spliced genes.
A:Reference number: A60238; MUID:91184292
```

```
A:Accession: A60238
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-56:1177-1231 <EST>
A:Note: only portions of this 4.3 kilobase mRNA were sequenced
R:Day, A.J.; Ripoche, J.; Lyons, A.; McIntosh, B.; Harris, T.J.R.; Sim, R.B.
Biochem. J. 201-207, 1987
A:Title: Sequence analysis of a cDNA clone encoding the C-terminal end of human complement factor H.
A:Reference number: A54726; MUID:88025472
A:Accession: A54726
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-56:1177-1231 <EST>
A:Cross-references: DB:117517; NID:g180497; PIDN:AAA52016.1; PID:g180498
A:Note: parts of this sequence were determined by protein sequencing
R:Ripoche, J.; Day, A.J.; Willis, A.C.; Belt, K.T.; Campbell, R.D.; Sim, R.B.
Biochem. J. 201-207, 1987
A:Title: Partial characterization of human complement factor H by protein and cDNA sequencing.
A:Reference number: A61565; MUID:86188123
A:Accession: A61565
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-56:1177-1231 <EST>
R:Sim, R.B.; Disceplo, R.G.
Biochem. J. 201-207, 1987
A:Title: Purification and structural studies on the complement-system control protein factor H.
A:Reference number: A26505; MUID:83048213
A:Accession: A26505
A:Molecule type: protein
A:Residues: 1-56:1177-1231 <EST>
R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, J.
Biochemistry 31, 3626-3634, 1992
A:Title: Solution structure of the fifth repeat of factor H: A second example of the protein fold.
A:Reference number: A44551; MUID:92232649
A:Contents: annotation; NMR structure determination, residues 264-292
R:Norman, D.G.; Barlow, P.N.; Baron, M.; Day, A.J.; Sim, R.B.; Campbell, I.D.
J. Mol. Biol. 219, 717-725, 1991
A:Title: Three-dimensional structure of a complement control protein module in solution.
A:Reference number: A49224; MUID:91278097
A:Contents: annotation; NMR structure determination, residues 927-985
R:Estallier, C.; Kolstinen, V.; Schwaeble, W.; Dierich, M.P.; Weiss, E.H.
J. Immunol. 146, 3190-3196, 1991
A:Title: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a novel isoform.
A:Reference number: I56100; MUID:91201892
A:Accession: I72654
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1047-1231 <RES>
A:Cross-references: GB:M65294; NID:g183766; PIDN:AAA35948.1; PID:g183767
R:Carroll, J.A.; Bates, R.C.; Smith, A.I.; Tetoz, T.; Arellano, A.; Gordon, D.L.; Burt, R.
Biochim. Biophys. Acta 1289, 305-311, 1996
A:Title: Factor H co-purifies with thrombospondin isolated from platelet secretate.
A:Reference number: S66298; MUID:96205365
A:Accession: S66298
A>Status: preliminary
A:Molecule type: protein
A:Residues: 411-419:574-578,580-582 <CAR>
C:Comment: Factor H has also been found bound to cell membranes in an unknown manner
C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver, spleen, and kidney.
C:Genetics: <HF1>
A:Gene: GDB:HF1; HF
A:Cross-references: GDB:120041; OMIM:134370
A:Map position: lq32-lq32
C:Genetics: <HF2>
A:Gene: GDB:HF2; HF
A:Cross-references: GDB:129095
A:Map position: lq32-lq32
A:Note: the correspondence between the two loci and the sequences indicated is unclear.
C:Function:
A:Description: a cofactor in the inactivation of C3b by serine proteinase 1; also involved in the alternative complement pathway
A:Pathway: complement alternate pathway
```


A;Description: a cofactor in the inactivation of C3b by serine proteinase I; also increased the alternative complement pathway
 C;Pathway: complement alternate pathway
 C;Superfamily: complement factor H; Complement factor H repeat homology
 C;Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma
 F;1-18/Domain: signal sequence #status predicted <SIG>
 F;19-44/Product: complement factor H, short splice form #status experimental <HAT>
 F;21-80/Domain: complement factor H repeat homology <FH01>
 F;85-141/Domain: complement factor H repeat homology <FH02>
 F;146-205/Domain: complement factor H repeat homology <FH03>
 F;210-262/Domain: complement factor H repeat homology <FH04>
 F;246-248/Region: cell attachment (R-G-D) motif
 F;267-320/Domain: complement factor H repeat homology <FH05>
 F;325-385/Domain: complement factor H repeat homology <FH06>
 F;389-442/Domain: complement factor H repeat homology <FH07>
 F;21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357-385/Binding site: carbohydrate (Asn) (covalent) #status absent
 F;217/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 69.6%; Score 1709; DB 1; Length 449;
Best Local Similarity 67.4%; Pred. No. 6.5e-112;
Matches 288; Conservative 47; Mismatches 92; Indels 0; Gaps 0;

[illegible]

RESULTS

S63551
factor H - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S65551
C:Reamers, C.J.; Day, A.J.; Sim, R.B.
Biochem. J. 315, 523-531, 1996
A:Title: Prediction from sequence comparisons of residues of factor H involved in the regulation of the alternative pathway of complement
A:Reference number: S65551 MUID:96202005

A;Accession: S00501
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-669 <SOA>

A/Cross-references: GB:X98697; NID:g1419423; PIDN:CA67257.1; PID:g1419424
C/Superfamily: complement factor H; complement factor H repeat homology
F:55-114/Domain: complement factor H repeat homology <FHR1>

F; 296-349/Domain: complement factor H repeat homology <FHR2>
F; 355-412/Domain: complement factor H repeat homology <FHR3>
F; 416-471/Domain: complement factor H repeat homology <FHR4>
F; 476-530/Domain: complement factor H repeat homology <FHR>
F; 538-592/Domain: complement factor H repeat homology <FHR5>
F; 599-651/Domain: complement factor H repeat homology <FHR6>

Query Match 50.6%; Score 1242; DB 2; Length 669;
Best Local Similarity 58.4%; Pred. NO. 3.8e-79;
Matches 206; Conservative 57; Mismatches 88; Indels 2; Gaps

Query Match 50.6%; Score 1242; DB 2; Length 669;
Best Local Similarity 58.4%; Pred. No. 3.8e-79;

Qy 76 GSFR LAGSEFEFGAKVYVTCDEGYQLGEIDYRECDADGWTNDIPICEVVKCLPVTELE 135

QY 136 NGRIVGAAEPDQEEYFQGVVRFECNSGFKIEGQEMHCSENGLWSNEKPOCWEISCLPP 195

Qy 196 RVENGDIYLPVYKENERFQYKCKQGFVYKERGDAVCTGSGWNPQSPCEEMTCLTPYIP 2555

Db 123 VILNGQAVLPKATYKQNERVQYRCAAGFEYQGRGDTVCTKSGWTPAPTCTEITCDPPRIP 182

Db 183 NGVYRPELSKYRGQDKITYECKKGFEEIRGTDTCTRDGWVPVPCAWKPCSYPIVKG 242

QY 316 RLYEESRRPFPVPICKEYSYCCDNGFTTPSQSWDYLRCTVNGWEPEVPCLRQCIFYH 375

QY 376 VEYGESSYQWRRYIEGQSAKVQCHSYSLPNGQDTYYCTENGWSPPPKCVRIK 428

Db 301 LENGTHHREEKYLOGETVRVHCYCSLQNDQNTWCTESGWSPPRCIRVK 353

RESULT 5

A35068
complement factor H-related protein 3A4/5G4 - mouse (fragment)
C.Species: Mus musculus (house mouse)

C. Speer, J. Mus. Nascim. (house mouse).
 C. Date: 27-Jul-1990 #sequence_revision 31-Dec-1993 #text_change 20-Aug-1999
 C. Accession: A35068; B35068; C35068; D35068; E35068; F35068; G35068
 R: Vik, D. P.; Munoz-Canoves, P.; Korzono, H.; Martin, L. G.; Tack, B. F.; Chapl.
 J. Biol. Chem. 265, 3193-3201, 1990

R;Vik, D.P.; Munoz-Canoves, P.; Kozono, H.; Martin, L.G.; Tack, B.F.; Chaplin, D.D.
J. Biol. Chem. 265, 3193-3201, 1990

A;Title: Identification and sequence analysis of four complement factor H-related trypsinogen-like proteases from *Paramecium* spp.
A;Reference number: A35070; MUID:90153969

A; Accession: A35068

A;Molecule type:- mRNA

A;ResIdues: 1-452 <VIK>

A:Cross-references: GB:M29010: NTD:σ192561: PTNN:AAA37A15 1: PTD:σ387128

A: Note: translation of the nucleotide sequence is not complete

A, note: translation of the nucleotide sequence is not complete

с, supplement: complement factor H; complement factor H repeat homology

F;28-81/Domain: complement factor H repeat homology <FHL>

F;86-146/Domain: complement factor H repeat homology <FH2>

F;150-203/Domain: complement factor H repeat homology <FH3>

F:209-266/Domain: complement factor H repeat homology <FH4>

F:270-325/Domain: complement factor H repeat homolog <EH5-

E:332-386/Domain: complement factor II repeat homology

Внутренний паспорт

F;390-431/Domain: complement factor H repeat homology <FH7>

Query Match 33.18; Score 813; PB 2; Length 452;

Best Local Similarity 71.08: Pred. No. 2.1e-49

Matches	137:	Conservative	18:	Mismatches	38:	Indels	0:	Gaps	0:
---------	------	--------------	-----	------------	-----	--------	----	------	----

QY 236 SGWNPOPSCEEMTCLTPYIPNGIYTPHRIKHRIDDEIRYECKNGFYPATRSPVSKCTITG 295

DB	15	TAWLSTAKGEEKTCSPYILNGIYTPHRIIHRSDDEIRYECNYGFYPTGTVSKCTPTG	74
QV	296	WIPAPRCSLKPCDEFQFKHGRLYVEESRRPYFPVPIGKEYSYVCDNGETTPSOSYWDYILR	355

Db 75 WIPVPRCTLKPCFEFPQFKYGRLYEESLRPNFPVPSIGNKYSRCDNGFSPSPSGYSWDYLR 134

QY 356 CTVNGWEPEVPCLCIRHVEYGESSYWRRIYEGOSAKVOCHSGYSLPNGQDTYYCTE 415
DB 135 CTAQGEPEVPCVRKCVFHYVENGSAWYWEKIYVQGLKVCYNGSYSLONGQDINTCTE 194
QY 416 NGWSPPPKCVRIK 428
DB 195 NGWSPPPKCVRIK 207

RESULT 6

apolipoprotein H-related protein 23L1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 27-Jul-1990 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
C:Accession: H35069; B35069; I35068
R:Vik, D.P.; Munoz-Canoves, P.; Kozono, H.; Martin, L.G.; Tack, B.F.; Chaplin, D.D.
J. Biol. Chem. 265, 3193-3201, 1990
A:Title: Identification and sequence analysis of four complement factor H-related trans
A:Reference number: A35070; MUID:90153969
A:Accession: H35068
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-303 <VIK>
A:Cross-references: GB:M29007; NID:g192557; PIDN:AAA37413.1; PID:g309164; GB:J05259
A:Note: translation of the nucleotide sequence is not complete
C:Superfamily: complement factor H repeat homology
C:Keywords: duplication
F:28-81/Domain: complement factor H repeat homology <FH01>
F:86-146/Domain: complement factor H repeat homology <FH02>
F:150-203/Domain: complement factor H repeat homology <FH03>
F:212-266/Domain: complement factor H repeat homology <FH04>

Query Match 33.0%; Score 811; DB 2; Length 303;
Best Local Similarity 70.8%; Pred. No. 1.9e-49;
Matches 136; Conservative 19; Mismatches 37; Indels 0; Gaps 0;

QY 236 SCWNQPSCEEMTCLTPYIPNGIYTPHRIKHRIIDDEIRVECKNGFYPATRSPVSKCTING 295
DB 15 TAWLSTAGEEKTCSPPILNIGIYTPHRIIHKSDDEIRTECKNGFYPVGTSVSKCTPTG 74
QY 296 WIPAPRCLSKPCDFPOFKHGRLYEESRRPYPVPPIGKEYSYCDNGFTTPSQSYWDYLR 355
DB 75 WIPVPRCTLKPCFEFFQYGRLYEESLRPNPVSIGNKYSYKCDNGFSPGSGYSDYLR 134
QY 356 CTVNGWEPEVPCLCIRHVEYGESSYWRRIYEGOSAKVOCHSGYSLPNGQDTYYCTE 415
DB 135 CTAQGEPEVPCVRKCVFHYVENGSAWYWEKIYVQGLKVCYNGSYSLONGQDINTCTE 194
QY 416 NGWSPPPKCVRI 427
DB 195 NGWSPPPKCVRI 206

RESULT 7

S46199
probable complement regulatory plasma protein SBI - barred sand bass
C:Species: Paralabrax nebulifer
C:Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 18-Jul-2001
C:Accession: S46199; S77894
R:Dahmen, A.; Kaidoh, T.; Zipfel, P.F.; Gigli, I.
Biochem. J. 301, 391-397, 1994
A:Title: Cloning and characterization of a cDNA representing a putative complement-regul
A:Reference number: S46199; MUID:94318039
A:Accession: S46199
A:Molecule type: mRNA
A:Residues: 1-1053 <DAHL>
A:Cross-references: EMBL:L21703; NID:g639894; PIDN:AAA92556.1; PID:g639895
A:Experimental source: liver
A:Accession: S77894
A:Molecule type: protein
A:Residues: 526-532, 'X', 534-537; 809-817, 'X', 819-826 <DAH2>

C:Genetics:

A:Gene: SBI
C:Superfamily: complement factor H repeat homology
C:Keywords: glycoprotein
F:89-145/Domain: complement factor H repeat homology <FH01>
F:334-389/Domain: complement factor H repeat homology <FH02>
F:450-502/Domain: complement factor H repeat homology <FH01>
F:569-624/Domain: complement factor H repeat homology <FH03>
F:682-738/Domain: complement factor H repeat homology <FH04>
F:743-802/Domain: complement factor H repeat homology <FH05>
F:935-989/Domain: complement factor H repeat homology <FH05>
F:993-1052/Domain: complement factor H repeat homology <FH06>

Query Match 23.2%; Score 569.5; DB 2; Length 1053;
Best Local Similarity 26.7%; Pred. No. 5.1e-32;
Matches 148; Conservative 60; Mismatches 169; Indels 177; Gaps 19;

QY 21 EQLYSGTQATYKCRPGYRTLGTIVKCKNGEMVPSNPSRICRKRPGHDPGDPFGSRL 80
DB 47 EASYPGGRQVRVCCNVGYS--GFFKLVCVEGRKWTGAK--CQPRSCGHPGDAQFADFLH 102
QY 81 AVGSEFEFGAKVYVTCDEGYQLLEIDYRECDADGNTNDIPICEVYVKLPVTELENGRIV 140
DB 103 AEGNDVFGSKVYVTCQKGYQWVSRYRRCVAGWDGVVPVCESSQC--PLIHVDNNVQV 161
QY 141 SGAAEPDQYYFGQVVRFCNCGFKI--EQKEMHCSENGLSNEKPOCVSEICLPPRVEN 199
DB 162 IGGPE---EATFGVVRFSKSRSEILDGSPELYCDERGDWSPGVPKCAITCAIPIEN 218
QY 200 GGIYLLKPVYKENERQYKCKGQVYKRGDAVCTGSG---WNPOPSCEEMTCL----- 250
DB 219 GNVPGAIREYKENDVLHYECDRAFKHIDR--PSTCIKQIKAEWSPTPLCESIKRLTIMD 277
QY 251 -TPYIP-----NGIYT----- 260
DB 278 GTRYEPAYRNLFSPGETLKVICARTSWISTPQTSVYVTTCDNGENSIPTCOEVRCSNR 337
QY 261 -PHRI-----KHRIDDEIRVECKNGFYPATRSPVSKCTITGWIPAP----- 300
DB 338 RPEHVDSDVRSWERYTLDNDTRYCKRGYKRTGGVTWATCGRNGWNPPLCEVKTSCKE 397
QY 301 -----RCSLKPCDFPOFKH 314
DB 398 NTQDAVIVGTQKIYNLNLKAIYACGEGNRGRITLTCGENGWSGDRKCTVKPCPLPKDP 457
QY 315 GRLYEESRRPYPVPPIGKEYSYCDNGFTTPSQSYWDYLRCTVNGWEPEVPCLCROI-- 372
DB 458 NGFF---RGPY---TGRVLYYTCQKGYKLFTEGHWAEAKCVDGVW---PELTTCISN 505
QY 373 -----FHYVEYGESSYWRRIYEGOSAKVOCHSGYSLPNGQDTYYCTENGW----- 418
DB 506 TTCGKFPPEIPNAEVI---RRYPEVQTVQVICNQGS--TQANSFSCENGWLLYGLSPDQ 560
QY 419 -----SPPPK 423
DB 561 ICTLRADVCGGPPPE 574

RESULT 8

D35069
complement factor H-related protein 9C4 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 27-Jul-1990 #sequence_revision 31-Dec-1993 #text_change 20-Aug-1999
C:Accession: D35069; B35070; C35070; D35070; E35069; F35070; F35069; F35070;
R:Vik, D.P.; Munoz-Canoves, P.; Kozono, H.; Martin, L.G.; Tack, B.F.; Chaplin, D.D.
J. Biol. Chem. 265, 3193-3201, 1990
A:Title: Identification and sequence analysis of four complement factor H-related tra
A:Reference number: A35070; MUID:90153969
A:Accession: D35069
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-808 <VIK>

A:Cross-references: GB:M29009; NID:q192563; PIDN:AAA37416.1; PID:g309166; GB:J05259

A:Note: translation of the nucleotide sequence is not complete

C:Superfamily: complement factor H; complement factor H repeat homology

F:28-88/Domain: complement factor H repeat homology <FH01>

F:92-145/Domain: complement factor H repeat homology <FH02>

F:151-208/Domain: complement factor H repeat homology <FH03>

F:212-267/Domain: complement factor H repeat homology <FH04>

F:272-325/Domain: complement factor H repeat homology <FH05>

F:332-386/Domain: complement factor H repeat homology <FH06>

F:393-446/Domain: complement factor H repeat homology <FH07>

F:455-505/Domain: complement factor H repeat homology <FH08>

F:511-564/Domain: complement factor H repeat homology <FH09>

F:569-622/Domain: complement factor H repeat homology <FH10>

F:627-681/Domain: complement factor H repeat homology <FH11>

F:688-742/Domain: complement factor H repeat homology <FH12>

F:746-807/Domain: complement factor H repeat homology <FH13>

Query Match 23.1%; Score 566; DB 2; Length 808;

Best Local Similarity 64.5%; Pred. No. 6.8e-32;

Matches 91; Conservative 23; Mismatches 27; Indels 0; Gaps 0;

QY 288 VSKCTITGIPAPRCSLKPCDFPQFKHGRLYEESRRPYFPVPIGKEYSYCDNGFTTFS 347

DB 9 LSNLLTAMLSAKGEVKSCPEPQFKYGRLYFEILLRPNFPVPSIGNKYSKCDNGESPPS 68

QY 348 QSYWDYLRCTVNGWEVPCRLQCIFHYVEYGESSYQWRYYIEGQSAKVOCHSGYSLPNG 407

DB 69 GLFWDYLRCTVNGWKPEVPCVRKCPHYVENGFEYWEKYYQGSQKLVQCYNGYSLQNG 128

QY 408 QDTYYCTENGWSPPKCVRIK 428

DB 129 QDTWCTENGWSPPKCIRIK 149

RESULT 9

S53711

C4BP alpha chain precursor - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: S53711

R:de Frutos, P.G.; Dahlbaeck, B.

Biochim. Biophys. Acta 1261, 285-289, 1995

A:Title: cDNA structure of rabbit C4b-binding protein alpha-chain. Preserved sequence mo

A:Reference number: S53711; MUID:95226458

A:Accession: S53711

A>Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-597 <DEF>

A:Cross-references: EMBL:235490

C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology

F:50-107/Domain: complement factor H repeat homology <FH1>

F:112-169/Domain: complement factor H repeat homology <FH2>

F:174-234/Domain: complement factor H repeat homology <FH3>

F:239-294/Domain: complement factor H repeat homology <FH4>

F:299-360/Domain: complement factor H repeat homology <FH5>

F:364-423/Domain: complement factor H repeat homology <FH6>

F:426-480/Domain: complement factor H repeat homology <FH7>

F:484-538/Domain: complement factor H repeat homology <FH8>

Query Match

Best Local Similarity 21.6%; Score 529.5; DB 1; Length 597;

Matches 148; Conservative 63; Mismatches 187; Indels 117; Gaps 26;

QY 2 DCKGPPPRENSIELSGSWSEQLYSEGTQATYKCRPGYRTLG-TIVKVK-NGEYVPSNPS 59

DB 49 DC-GPPHLLFASSISELSENEYQGTILKYTCRPGYTRNGLNPILTCRPLGLM---SYD 104

QY .60 RICRRPCGHPGCDTFFGFRFRLAGSEFEFGAKVVTYDCGQVOLLGEIDYRECDAD---G 115

DB 105 TFCVKKRCRNPGLDPLNG--QVEVKTDFFSGQIEFSCSEGYILLIGSTT-SHCDIQEKGV 161

QY 116 WTNDIPICEVVKCLPVTLENGRIVSGAAEPDQYFYFGVYVFECSGFKIEGQKEMHCS 175

DB 162 WSDPLKCEIVKCEPPNIIINGKHNG---NEDIHTYGSVVTYSCNPRFSLLEASISCT 218

QY 176 EN---GLASNEKPOCVESICLPPRVENGDIY-LAPVYKENERFQYKCKQGFYKERGD 230

DB 219 VKNKTGVWSPSPVCKEIIICSPNPVPHKIIISGFGPIYNYKDSIMYTCIDGFLVRGSSL 278

QY 231 AVC-TGSGWNQPS-CEEMTCL-TPYIPN-----GIYTPHRIKHRIDDEIRYE 275

DB 279 IHCELDKSNPSPVCESSNCLGLPNVPHASOQGYOWSTKEGYVS-----VCTELRYK 331

QY 276 CKNGFYPATRSPVSKCTIT-----GWIPAPRCSLKPCDFPQFKHGRLYYEESRRP----- 325

DB 332 CRPGYRPADEPI---IVTCQEDLRWSPAGCEAICCPHPQLDNGAI-TEHSRNPANNC 387

QY 326 -YF-----PVPIG----- 332

DB 388 TYFEGDIVVYCYQKQKORTTTCRADGTWSPHTPLCGSPCDSPPIAHGYHTPVSLSFT 447

QY 333 KEYSYCDNGFTTSPQSYWDYLRCTVNGWEVPC-CLRQCIFHYVEYGESSYQWRYYIEG 391

DB 448 QEVEYDCEEGYTLVGER---KLSGSSSINSKAPPOCKALCPKPEITNGKLSVYVKAQYVEK 504

QY 392 QSAKVOCHSGYSLPNQDQTYCYCTEN-GWSP-PC 424

DB 505 ENLTVCRCDSGYGVV-GSPSITCTENRNWQPEVPC 538

RESULT 10

NBHUC4

C4b-binding protein alpha chain precursor - human

N:Alternate names: C4BP; proline-rich protein

C:Species: Homo sapiens (man)

C:Date: 13-Aug-1986 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

C:Accession: A33568; S02372; A90326; A24182; A93134; S29492; A31785; I52244; A03210

R:Watsuguchi, T.; Okamura, S.; Asao, T.; Sata, T.; Niho, Y.

Biochem. Biophys. Res. Commun. 165, 138-144, 1989

A:Title: Molecular cloning of the cDNA coding for proline-rich protein (PRP): Identifi

A:Reference number: A33568; MUID:90073699

A:Accession: A33568

A:Molecule type: mRNA

A:Residues: 1-597 <NAI>

A:Cross-references: GB:IM31452; NID:q190501; PIDN:AAA36507.1; PID:gl90502

A:Note: the authors translated the codon GGA for residue 492 as Glu

R:Lintin, S.J.; Lewin, A.R.; Reid, K.B.M.

FEBS Lett. 232, 328-332, 1988

A:Title: Derivation of the sequence of the signal peptide in human C4b-binding prote

A:Reference number: S02372; MUID:86242821

A:Accession: S02372

A:Molecule type: mRNA

A:Residues: 17-81 <LI2>

A:Cross-references: EMBL:X07853

A:Note: although the sequence determined extends to residue 9 above, these authors a

R:Chung, L.P.; Bentley, D.R.; Reid, K.B.M.

Biochem. J. 230, 133-141, 1985

A:Title: Molecular cloning and characterization of the cDNA coding for C4b-binding F

A:Reference number: A90326; MUID:86025405

A:Accession: A90326

A:Molecule type: mRNA

A:Residues: 80-597 <CH2>

A:Cross-references: GB:X02865; NID:g29564; PIDN:CAA26617.1; PID:g29565

R:Lintin, S.J.; Reid, K.B.M.

FEBS Lett. 204, 77-81, 1986

A:Title: Studies on the structure of the human C4b-binding protein gene.

A:Reference number: A24182; MUID:86301119

A:Accession: A24182

A:Molecule type: DNA

A:Residues: 203-288 <LIN>

A:Cross-references: EMBL:X04284; EMBL:X04296

R:Rodriguez de Cordoba, S.; Sanchez-Corral, P.; Rey-Campos, J.

J. Exp. Med. 173, 1073-1082, 1991

A::Title: Structure of the gene coding for the alpha polypeptide chain of the human complement C4B-binding protein.

A::Reference number: A43023; PMID:91217619

A::Contents: annotation: exon-intron boundaries

R::Chung, L.P.; Gagnon, J.; Reid, K.B.M.
Mol Immunol. 22: 427-435, 1985

A::Title: Amino acid sequence studies of human C4b-binding protein: N-terminal sequence determined by cyanogen bromide treatment.

A::Reference number: A93134; PMID:85296001

A::Accession: A93134

A::Molecule type: protein

A::Residues: 48-81 <CH1>

A::Note: This paper reports amino-terminal sequences of the intact protein and of a number of fragments generated by cyanogen bromide treatment.

R::Hessing, M.; Kanters, D.; Takeya, H.; van't Veer, C.; Hackeng, T.M.; Iwanaga, S.; Bommberg, R.
FEBS Lett. 317: 228-232, 1993

A::Title: The region Ser(333)-Arg(356) of the alpha-chain of human C4b-binding protein is essential for its binding to C4b.

A::Reference number: S29492; PMID:93146164

A::Accession: S29492

A::Status: preliminary

A::Molecule type: protein

A::Residues: 381-404 <HES>

R::Suzuki, K.; Nishio, J.
J Biol Chem. 263: 17034-17039, 1988

A::Title: Binding site for vitamin K-dependent protein S on complement C4b-binding protein

A::Reference number: A31785; PMID:89034204

A::Accession: A31785

A::Molecule type: protein

A::Residues: 495-505,'X',507-510,'X',512-515 <SUZ>

A::Note: this peptide appears to bind protein S

R::Dahlback, B.; Smith, C.A.; Muller-Eberhard, H.J.
Proc Natl Acad Sci U.S.A. 80: 3461-3465, 1983

A::Title: Visualization of human C4b-binding protein and its complexes with vitamin K-dependent protein S

A::Reference number: A93950; PMID:83221615

A::Contents: annotation; electron microscopy: three-dimensional structure; ligand binding

R::Asou, T.; Okamura, S.; Matsuguchi, T.; Sakamoto, N.; Sata, T.; Niho, Y.
Biochem Biophys Res Commun. 174: 222-227, 1991

A::Title: Genomic organization of the alpha chain of the human C4b-binding protein gene.

A::Reference number: I52244; PMID:91113199

A::Accession: I52244

A::Status: translation not shown; translated from GB/EMBL/DDBBY

A::Molecule type: DNA

A::Residues: 1-597 <ASO>

A::Cross-references: GB:M62486; NID:g190498; PIDN:AAA36506.1; PID:g190500

C::Comment: C4BP controls the classical pathway of complement activation. It binds as a dimer to the degradation of the C4C2a complex (C3 convertase) by dissociating the complement fragment C4b from C2a. The beta chain binds the vitamin K-dependent plasma protein S. A minor isoform of C4BP occurs in plasma in two forms, both of which bind complement fragment C4b via disulfide bonds. The beta chain binds the vitamin K-dependent plasma protein S. A minor isoform of C4BP occurs in plasma in two forms, both of which bind complement fragment C4b via disulfide bonds.

C::Comment: The molecule has a central body supporting seven tentacles (alpha chains), each containing a C4b-binding site.

A::Gene: C4BPA

A::Cross-references: GDB:I20568; OMIM:120830

A::Map position: lq32-lq32

A::Introns: 48/1; 110/1; 143/2; 236/1; 297/1; 362/1; 425/1; 482/1; 540/3

C::Complex: octamer of seven alpha chains and one beta chain

C::Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology

C::Keywords: acute phase; chylomicron; complement pathway; duplication; glycoprotein; plasminogen activator; signal sequence

F::1-48/Domain: signal sequence #status predicted <SIG>

F::49-597/Product: C4b-binding protein alpha chain #status predicted <MAT>

F::50-108/Domain: complement factor H repeat homology <PH1>

F::113-170/Domain: complement factor H repeat homology <PH2>

F::175-234/Domain: complement factor H repeat homology <PH3>

F::239-294/Domain: complement factor H repeat homology <PH4>

F::299-360/Domain: complement factor H repeat homology <PH5>

F::364-422/Domain: complement factor H repeat homology <PH6>

F::381-404/Region: complement C4b binding #status predicted

F::426-480/Domain: complement factor H repeat homology <PH7>

F::484-538/Domain: complement factor H repeat homology <PH8>

F::221,506,528/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 19.6%; Score 481.5; DB 1; Length 597;
Best Local Similarity 26.5%; Pred. No. 3.8e-26;
Matches 134; Conservative 78; Mismatches 195; Indels 99; Gaps 25;

Db 331 CEVKSDAIPNHLGRVFLPPN-----LQLGAEVSEFVCDLGLQKGRKPS-----QC 378
QY 357 TVNG-----WEPEVPCLRQCIFH---YVEYGESSYWRRIEIQSAKVOCHSCYSPLNGCD 409
Db 379 IPEGETVIVNNRFPVCEQISCDDPPPEVKNARKPYISLPIVPGTVLRYTCSPSYRL-IGEK 437
QY 410 TYVC-TEN---GW-SPPKK 424
Db 438 AIFCISENQVHATWDKAPIC 458
RESULT 14
S57953
C4BP protein alpha chain precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 20-Aug-1999
C:Accession: S57953
R: Hillarp, A.; Thern, A.; Dahlback, B.
submitted to the EMBL Data Library, July 1995
A:Description: Molecular cloning of rat C4b-binding protein alpha- and beta-chains: structure and function
A:Reference number: S57953
A:Accession: S57953
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-558 <HIL>
A:Cross-references: EMBL:250051; NID:q899379; PIDN:CAA90391.1; PID:q899380
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
F:15-72/Domain: complement factor H repeat homology <FH1>
F:139-199/Domain: complement factor H repeat homology <FH2>
F:204-238/Domain: complement factor H repeat homology <FH3>
F:263-324/Domain: complement factor H repeat homology <FH4>
F:328-386/Domain: complement factor H repeat homology <FH5>
F:390-443/Domain: complement factor H repeat homology <FH6>
F:447-501/Domain: complement factor H repeat homology <FH8>
Query Match 18.3%; Score 449; DB 2; Length 558;
Best Local Similarity 26.7%; Pred. No. 6.5e-24;
Matches 137; Conservative 71; Mismatches 198; Indels 108; Gaps 24;
QY 5 GPPRENSILLGSWSEQLSEGTQATYKCRGY-RTLGTVKVKCN-GEWVSPNSRIC 62
Db 16 GPPDLPVLPALPASENNQDFFSHHTTLRYNCRGYSRASSOSLYCKPLGKW---QINAC 72
QY 63 RKRPGHDPDPFGSFRVLAVSEFEFGAKVYVTCDEGYQLLG-EIDYRECDAG--WTND 119
Db 73 VKKSCRNPGLDQNG--KVEVXTDLFLGQIEFSCSEGVILIGSSTSYCEIQKGYSWSDP 130
QY 120 IPICEVVKCLPVTELENGRIVSGAAEPDQYVYFGVVRFECSNGPKIEGKEMHCS---- 175
Db 131 LPECVIAKCGMPDPTISNGK---HNGREBEFFTYRSVYTKCDPDFTLLGNASITCTVYVK 187
QY 176 ENGLWSNEKPOCVETISCLPPRVENGD-GIYKLPVYKENERQYKCKQGVYKERCDAVCT 234
Db 188 TVGWSNPSPPTCERIIICWPVKVLTHTINSRGKHYTKYKDSVRFCQKGFVLGRSGVHICE 247
QY 235 GSG-WNPQSCSEMTCL-TPYIPNG--IYTPHRIK---HRIDDEIRYECKNGFYPATRSP 287
Db 248 ADGWSNPVPCVCLNSCTDIPDIPNAALITSPRKEDYVPGTVLRYICRPGYEPATRQP 307
QY 288 VSKCTI-----TGWIAPRCSLKCPDFQFKHRLYYEESRRP-----YFPVPIGKEYSY 337
Db 308 M---TVICQKDLWSMLRGCKEICPCVPDPKPSVRVQIEKAHPNDCTYF---FGDEVSY 361
QY 338 YCDN-----GFTTPSQSYWDYLR----- 355
Db 362 TCQNDIMLTATKSDGTWHPTPTPSCHOSCDPFPALAHGRYTKSSY---YVRQTVYECEE 419
QY 356 -----CTVNGWPEVP-CLRQICIFHYVEGESSYWRRIEIQSAKVOCHSCYS 403
Db 420 GYRLVGENTISCWTSQWTPAPQCKALCRKPEIGNGVLSWTNKOQTVETENVTIOCDSGVF 479

QY 404 LPNGODTYVYCTENG-WSP-----PPKCVRI 427
Db 480 M-LGQSITCSENGTWYPKVSRCEQEVPRKCEHV 512
RESULT 15
I36936
complement receptor 1 - chimpanzee (fragment)
C:Species: Pan troglodytes (chimpanzee)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C:Accession: I36936; I36937
J: Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.
J. Immunol. 153, 691-700, 1994
A:Title: Primary sequence of an alternatively spliced form of CRI. Candidate for the
A:Reference number: I36935; MUID:94292799
A:Accession: I36936
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2014 <BIR1>
A:Cross-references: GB:L24920; NID:g551564; PIDN:AAA51438.1; PID:g557725
A:Accession: I36937
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-397,1751-2014 <BIR2>
A:Cross-references: GB:L24921; NID:g557726; PIDN:AAA51439.1; PID:g557727
C:Genetics: CRI
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C:Keywords: duplication; glycoprotein; tandem repeat
F:18-74/Domain: complement factor H repeat homology <FH01>
F:79-136/Domain: complement factor H repeat homology <FH02>
F:141-207/Domain: complement factor H repeat homology <FH03>
F:213-268/Domain: complement factor H repeat homology <FH04>
F:272-338/Domain: complement factor H repeat homology <FH05>
F:333-391/Domain: complement factor H repeat homology <FH06>
F:396-462/Domain: complement factor H repeat homology <FH07>
F:470-526/Domain: complement factor H repeat homology <FH08>
F:529-586/Domain: complement factor H repeat homology <FH09>
F:723-778/Domain: complement factor H repeat homology <FH10>
F:1041-1107/Domain: complement factor H repeat homology <FH11>
F:1172-1228/Domain: complement factor H repeat homology <FH12>
F:1233-1291/Domain: complement factor H repeat homology <FH13>
F:1296-1362/Domain: complement factor H repeat homology <FH14>
F:1432-1489/Domain: complement factor H repeat homology <FH15>
F:1625-1681/Domain: complement factor H repeat homology <FH16>
F:1749-1815/Domain: complement factor H repeat homology <FH17>
F:1823-1879/Domain: complement factor H repeat homology <FH18>
F:1884-1940/Domain: complement factor H repeat homology <FH19>
Query Match 18.2%; Score 447; DB 2; Length 2014;
Best Local Similarity 30.2%; Pred. No. 3.7e-23;
Matches 137; Conservative 64; Mismatches 165; Indels 88; Gaps 32;
QY 27 GTQATYKCRPGYRTLGTVKVKCNKEWVSPNSRICRKRCPGDPDPFGSFRLVAGSFR 86
Db 941 GTSKYECRPEYGRPFSTICLDNLVW--SSPKVCRKSKCTPPDPVNG--MVHVIIDI 996
QY 87 EFGAKVYTCDEGYQLLGIEDYRECDAG----WTNDIPICEVVKC-LPVTELENGRIVS 141
Db 997 QVGSRIYSCCTGHRLLIGH-SSAECILSGNSAHWSTKPKICQRIPCGLPPT-IANGDFIS 1054
QY 142 GAAEPDQYVYFGVVRFECSNG-----FKTEGKEMHCSN-----GLWSNEKPOC-VEI 190
Db 1055 TNRE---NFHYGSVVYTCNLGSRGRKVFELVGEPSIYCTSNDDQGVWSGAPQCIIPN 1111
QY 191 SCLPPRVENG---DGIYKLPVYKENERQYKCKQGVYKERCDAVCTG-SGNPQ-PSC 244
Db 1112 KCTPNVNGILVSDN----RSLFSLNVEVFRCQGFVYKGRVVKVKCALNKEPELPS 1168
QY 245 EEMTCLTPYIPNGIYTP-HRIKHRIDDEIRYECKNGFYPATRSPVSKCTITG-WTP-APR 301

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:12:05 ; Search time 21.93 Seconds
(without alignments)
755.676 Million cell updates/sec

Title: US-09-316-163-14

Perfect score: 2454

Sequence: 1 EDCKGPPPPRENSILSGWS.....DTYYCTENGWSPPPKCVRIK 428

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2055	83.7	1234	1	CPAH_MOUSE
2	1714	69.8	1231	1	CPAH_HUMAN
3	1272.5	51.9	685	1	CPAH_BOVIN
4	481.5	19.6	597	1	C4BP_HUMAN
5	478	19.5	330	1	FHR3_HUMAN
6	474.5	19.3	610	1	C4BP_BOVIN
7	449	18.3	558	1	C4BP_RAT
8	437	17.8	2039	1	CR1_HUMAN
9	384.5	15.7	469	1	C4BP_MOUSE
10	373	15.2	1025	1	CR2_MOUSE
11	368	15.0	1033	1	CR2_HUMAN
12	336	13.7	661	1	F13B_HUMAN
13	332.5	13.5	331	1	FHR4_HUMAN
14	332	13.5	668	1	F13B_MOUSE
15	315	12.8	330	1	FHR1_HUMAN
16	314.5	12.8	263	1	VCP_VACCV
17	310.5	12.7	768	1	LEM3_MOUSE
18	310	12.6	270	1	FHR2_HUMAN
19	309.5	12.6	830	1	LEM3_HUMAN
20	298.5	12.2	377	1	MCP_HUMAN
21	296	12.1	768	1	LEM3_RAT
22	290	11.8	610	1	LEM2_HUMAN
23	282.5	11.5	769	1	LEM3_SHEEP
24	277.5	11.3	611	1	LEM2_CANFA
25	272	11.1	612	1	LEM2_MOUSE
26	271.5	11.1	345	1	APOH_MOUSE
27	267.5	10.9	345	1	APOH_HUMAN
28	267.5	10.9	507	1	DAF_CAVPO
29	262.5	10.7	345	1	APOH_CANFA
30	256.5	10.5	551	1	LEM2_RABIT
31	255	10.4	360	1	CCPH_HSVSA
32	252	10.3	484	1	LEM2_PIG
33	246.5	10.0	646	1	LEM3_BOVIN

RESULT 1

ID	CPAH_MOUSE	STANDARD;	PRT: 1234 AA.
AC	P06909;		
DT	01-JAN-1988 (Rel. 06, Created)		
DT	01-JAN-1988 (Rel. 06, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Complement factor H precursor (Protein beta-1-H).		
GN	HFI1 OR CFH.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=86233353; PubMed=2940596;		
RA	Kristensen T., Tack B.F.;		
RT	"Murine protein H is comprised of 20 repeating units, 61 amino acids in length.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 83:3963-3967(1986).		
RN	[2]		
RP	SEQUENCE OF 1-19 FROM N.A.		
RC	STRAIN=BALB/C;		
RX	MEDLINE=90148935; PubMed=2533512;		
RA	Munoz-Canoves P., Tack B.F., Vik D.P.;		
RT	"Analysis of complement factor H mRNA expression: dexamethasone and IFN-gamma increase the level of H in L cells.";		
RL	Biochemistry 28:9891-9897(1989).		
RN	[3]		
RP	SEQUENCE OF 1-18 FROM N.A.		
RX	MEDLINE=90111033; PubMed=2136885;		
RA	Natsume-Sakai S., Nonaka M., Nonaka M., Harada Y.N., Shreffler D.C., Moriaki K.;		
RT	"Demonstration of an unusual allelic variation of mouse factor H by the complete cDNA sequence of the H.2 allotype.";		
RL	J. Immunol. 144:358-362(1990).		
CC	1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3B)NBB COMPLEX (C5 CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.		
CC	1- POLYMORPHISM: TWO CODOMINANT ALLELES OF FACTOR H ARE PRESENT IN MICE.		
CC	1- SIMILARITY: CONTAINS 20 SUSHI (SCR) DOMAINS.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	EMBL; M12660; AAA37759.1; -		
DR	EMBL; J02891; AAA37795.1; -		
DR	EMBL; M31979; AAA37762.1; -		
DR	PIR; A26154; NBMSH.		

P08174 homo sapien
Q61475 mus musculus
Q09101 drosophila
P49457 pongo pygma
P17690 bos taurus
P98105 rattus norv
Q61476 mus musculus
P98107 bos taurus
P26644 rattus norv
P18337 mus musculus
P28175 tachypneus
Q26422 carnoscor

ALIGNMENTS

FT	CARBOHYD	1030	1030	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1061	1061	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1225	1225	N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQ	SEQUENCE	1234 AA;	139082 MW;	CSAC02F341B957E7 CRC64;

Query Match	83.7%;	Score 2055;	DB 1;	Length 1234;
Best Local Similarity	82.9%;	Pred. No. 1.6e-155;		
Matches 355;	Conservative 28;	Mismatches 45;	Indels 0;	Gaps

QY	1	EDCKGPPPRENSEILSGSWSEQLYSEGTQATYKCRPGYRTLGTIVKVKNGEWPVNSPSR	60
DB	19	EDCKGPPPRENSEILSGSWSEQLYPEGTQATYKCRPGYRTLGTIVKVKNGEWPVNSPSR	78
QY	61	ICRKRRCPGHPGDPFGFSFRILAVGSEFEFGAKVYVYTCDDGYQLLGEIDYRECCDADGWTNDI	120
DB	79	ICRKKPCGHPGDPFGFSFRILAVGSPFEFGAKVYVYTCDDGYQLLGEIDYRECCDADGWTNDI	138
QY	121	PICBVKACLPVTELENGRIVSGAEPDOEYFYGVVRFECNSGFKIEGOKEMHCSENGLW	180
DB	139	PLCEBVKACLPVTELENGRIVSGAETDOEYFYGVVRFECNSGFKIEGHEIHCSENGLW	198
QY	181	SNEKPCQVEILCTPPRVENGDDGYLKPVYKENERFOYKCKQGSPVYKRGDVAUCTGSGWNP	240
DB	199	SNEKPCQVEILCTPPRVENGDDGYNKPVYKENERHYKCKHGKGVYKRGDVAUCTGSGWSS	258
QY	241	QPSCEEMTCLTPYIPNGIYTPHRIKHRIIDDEIRYECNKGFGYPATRSPVSKCTITGHIPAP	300
DB	259	QPFCEKRCSPPYTLNGIYTPHRIIHRSDDEIRYECNKGFGYPVGTGTVSKCTPTGTGWIPV	318
QY	301	RCSLKPCCDFQFKHGRLYVEESRRPFPVPPIGRKEYSYVCDNGFTTSPQSRYWDYLRCRTVNG	360
DB	319	RCTLKPCEFPQFKYGRLYVEESLRPNFPYSIGNKYSYKCDNGFSPSGYSWDYLRCRTAOG	378
QY	361	WEPEVPCRLCROCFHYHYEYSGSSYQWQRYIEGSAKYQCHSGSLPNQGDYTYCTENGWSP	420
DB	379	WEPEVPCVRKCVHYEYSGSSYQWQRYIEGSAKYQCHSGSLPNQGDYTYCTENGWSP	438
QY	421	PPKCVRIRK 428	
DB	439	PPKCIRIK 446	

RESULT 2	
CFAH_HUMAN	STANDARD: PRT; 1231 AA.
ID	CFAH_HUMAN
AC	P08603; Q14570; P78435; Q9NU86;
DT	01-AUG-1988 (Rel. 08, Created)
DT	01-JAN-1990 (Rel. 13, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Complement factor H precursor (H factor 1).
GN	HF1 OR HF OR CFH.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON	NCBI_TaxID=9606;
OX	[1]
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT Y-402.
RC	TISSUE=Liver;
RX	MEDLINE=88134059; PubMed=2963625;
RA	Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;
RT	"The complete amino acid sequence of human complement factor H.;"
RL	Biochem. J. 249:593-602(1988).
RL	[2]
RP	SEQUENCE OF 53-445 FROM N.A.
RX	MEDLINE=87054207; PubMed=2946589;
RA	Schulz T.F., Schwaebler W., Stanley K.K., Weiss E., Dierich M.P.;
RT	"Human complement factor H: Isolation of cDNA clones and partial cDNA
RT	sequence of the 38-kDa tryptic fragment containing the binding site
RT	for C3b.;"
RL	Eur. J. Immunol. 16:1351-1355(1986).
RL	[3]

RX MEDLINE-86169701; PubMed-2937845;
 RA Kristensen T., Wetsel R.A., Tack B.F.;
 RT "Structural analysis of human complement protein H: homology with C4b
 binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";
 RL J. Immunol. 136:3407-3411(1986).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RA Bird C.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 1047-1231 FROM N.A.
 RX MEDLINE-91201892; PubMed-1826708;
 RA Estaller C., Koistinen V., Schwaible W., Dierich M.P., Weiss E.H.;
 RT "Cloning of the 1.4-kb mRNA species of human complement factor H:
 reveals a novel member of the short consensus repeat family related
 to the carboxy terminal of the classical 150-kDa molecule.";
 RL J. Immunol. 146:3190-3196(1991).
 RN [6]
 RP SEQUENCE OF 19-35.
 RX MEDLINE-83048213; PubMed-6215918;
 RA Sim R.B., Discipio R.G.;
 RT "Purification and structural studies on the complement-system control
 protein beta 1H (factor H).";
 RL Biochem. J. 205:285-293(1982).
 RN [7]
 RP SEQUENCE OF 1-19 FROM N.A.
 RA Vik D.P., Williams S.A.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE OF 1-9 FROM N.A.
 RA Dominguez O.;
 RL Thesis (1993), Hospital Trias I Pujol, Spain.
 RN [9]
 RP STRUCTURE BY NMR OF 927-985 (SUSHI 16).
 RX MEDLINE-91278097; PubMed-1829116;
 RA Norman D.G., Barlow P.N., Baron M., Day A.J., Sim B., Campbell I.D.;
 RT "Three-dimensional structure of a complement control protein module
 in solution.";
 RL J. Mol. Biol. 219:717-725(1991).
 RN [10]
 RP STRUCTURE BY NMR OF 264-322 (SUSHI 5).
 RX MEDLINE-92232649; PubMed-1533152;
 RA Barlow P.N., Norman D.G., Steinkasserer A., Horne T.J., Pearce J.,
 RL Driscoll P.C., Sim B., Campbell I.D.;
 RT "Solution structure of the fifth repeat of factor H: a second example
 of the complement control protein module.";
 RL Biochemistry 31:3626-3634(1992).
 RN [11]
 RP STRUCTURE BY NMR OF 866-985 (SUSHI 15 AND 16).
 RX MEDLINE-93323119; PubMed-8331663;
 RA Barlow P.N., Steinkasserer A., Norman D.G., Kieffer B., Wiles A.P.,
 RL Sim B., Campbell I.D.;
 RT "Solution structure of a pair of complement modules by nuclear
 magnetic resonance.";
 RL J. Mol. Biol. 232:268-284(1993).
 CC -1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF
 C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE
 C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3B)NBB COMPLEX (C5
 CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
 PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: CONTAINS 20 SUSHI (SCR) DOMAINS.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION
 341 ONWARD DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Y00716; CAA68704.1; -;
 DR EMBL; X04697; CABA1739.1; ALT_FRAME.
 DR EMBL; X07523; CAA30403.1; -;
 DR EMBL; M12383; AAA52013.1; -;
 DR EMBL; AL049744; CAB70598.1; -;
 DR EMBL; M65294; AAA35948.1; -;
 DR EMBL; U56979; AAB01987.1; -;
 DR EMBL; Z29665; CAA82763.1; -;
 DR PIR; S00254; NBHUH.
 DR PIR; S03013; S03013.
 DR PDB; 1HCC; 15-APR-92.
 DR PDB; 1HPH; 15-JUL-93.
 DR PDB; 1HFI; 15-JUL-93.
 DR MIM; 134370; -;
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; sushi; 20.
 DR SMART: SM00032; CCP; 20.
 KW Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;
 KW Signal; 3D-structure; Polymorphism; Alternative splicing.
 FT CHAIN 19 1231 COMPLEMENT FACTOR H.
 FT SIGNAL 1 18
 FT DOMAIN 20 81 SUSHI 1.
 FT DOMAIN 84 142 SUSHI 2.
 FT DOMAIN 145 206 SUSHI 3.
 FT DOMAIN 209 263 SUSHI 4.
 FT DOMAIN 266 321 SUSHI 5.
 FT DOMAIN 324 386 SUSHI 6.
 FT DOMAIN 388 443 SUSHI 7.
 FT DOMAIN 447 506 SUSHI 8.
 FT DOMAIN 508 565 SUSHI 9.
 FT DOMAIN 568 624 SUSHI 10.
 FT DOMAIN 629 685 SUSHI 11.
 FT DOMAIN 690 745 SUSHI 12.
 FT DOMAIN 752 804 SUSHI 13.
 FT DOMAIN 810 865 SUSHI 14.
 FT DOMAIN 869 927 SUSHI 15.
 FT DOMAIN 930 985 SUSHI 16.
 FT DOMAIN 1047 1103 SUSHI 17.
 FT DOMAIN 1108 1164 SUSHI 18.
 FT DOMAIN 1166 1229 SUSHI 19.
 FT DISULFID 21 66 BY SIMILARITY.
 FT DISULFID 52 80 BY SIMILARITY.
 FT DISULFID 85 129 BY SIMILARITY.
 FT DISULFID 114 141 BY SIMILARITY.
 FT DISULFID 146 192 BY SIMILARITY.
 FT DISULFID 178 205 BY SIMILARITY.
 FT DISULFID 210 251 BY SIMILARITY.
 FT DISULFID 237 262 BY SIMILARITY.
 FT DISULFID 267 309 BY SIMILARITY.
 FT DISULFID 294 320 BY SIMILARITY.
 FT DISULFID 325 374 BY SIMILARITY.
 FT DISULFID 357 385 BY SIMILARITY.
 FT DISULFID 389 431 BY SIMILARITY.
 FT DISULFID 416 442 BY SIMILARITY.
 FT DISULFID 448 494 BY SIMILARITY.
 FT DISULFID 477 505 BY SIMILARITY.
 FT DISULFID 509 553 BY SIMILARITY.
 FT DISULFID 536 564 BY SIMILARITY.
 FT DISULFID 569 611 BY SIMILARITY.
 FT DISULFID 597 623 BY SIMILARITY.
 FT DISULFID 630 673 BY SIMILARITY.
 FT DISULFID 659 684 BY SIMILARITY.
 FT DISULFID 691 733 BY SIMILARITY.
 FT DISULFID 719 744 BY SIMILARITY.
 FT DISULFID 753 792 BY SIMILARITY.
 FT DISULFID 781 803 BY SIMILARITY.
 FT DISULFID 811 853 BY SIMILARITY.
 FT DISULFID 839 864 BY SIMILARITY.
 FT DISULFID 870 915 BY SIMILARITY.
 FT DISULFID 901 926 BY SIMILARITY.
 FT DISULFID 931 973 BY SIMILARITY.
 FT DISULFID 959 984 BY SIMILARITY.

```
FT DISULFID 989 1032 BY SIMILARITY.
FT DISULFID 1018 1043 BY SIMILARITY.
FT DISULFID 1048 1091 BY SIMILARITY.
FT DISULFID 1077 1102 BY SIMILARITY.
FT DISULFID 1109 1152 BY SIMILARITY.
FT DISULFID 1138 1163 BY SIMILARITY.
FT DISULFID 1167 1218 BY SIMILARITY.
FT DISULFID 1201 1228 BY SIMILARITY.
FT CARBOHYD 529 529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 718 718 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 822 822 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 882 882 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 911 911 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1029 1029 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1095 1095 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 446 449 KTCSS -> SFTL (IN ISOFORM 2).
FT VARSPLIC 450 1231 MISSING (IN ISOFORM 2).
FT VARIANT 402 402 H -> Y.
FT CONFLICT 21 21 /FTID-VAR_001979.
FT CONFLICT 30 30 C -> Q (IN REF. 3).
FT CONFLICT 34 34 T -> V (IN REF. 3).
FT CONFLICT 53 54 T -> Q (IN REF. 3).
FT STRAND 870 871 RP -> IL (IN REF. 2).
FT STRAND 876 876

Query Match 69.8%; Score 1714; DB 1; Length 1231;
Best Local Similarity 67.5%; Pred. No. 1.9e-128;
Matches 289; Conservative 47; Mismatches 92; Indels 0; Gaps 0;

QY 1 EDCCKPPPRENSILSGSSEQLYSRGTOATYKCRGYFTLTGYVCKNGEWPSPNPSR 60
Db 19 EDCNELPPRRNTEILGWSDDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGWVALNPLR 78
QY 61 ICRKRCGHPGDTFGSFRFLAVGSEPEFGAKVYVTCDEGYQLLGEIDYRECDADGWTNDI 120
Db 79 KCOKRCGHPGDTFGTFTLTGNGVFEYGVKAVYTCNEGYQLLGEINRECDTGDWTNDI 138
QY 121 PICEVVKCLPTELENGRIVSGAAEPDQYFYFGVVRFCNSGFKTEGQKHCSENGLMW 180
Db 139 PICEVVKCLPVTAPENGKIVSSAMEPDREYHFQAVRFVCSGYKTEGDEEMHCSDDGF 198
QY 181 SNEKPCVCEISCLPPRVENGDTLYLKVYKENERFOYKCKQGFVYKERGDAVCTGSGWNP 240
Db 199 SNEKPCVCEISCKSPDVINGSPISOKIYKENERFOYKCMNGEYSESGDAVCTGSGWRP 258
QY 241 QPSCEMTCLTPYIPNGIYTPHRIKHRIHDEIRYCKNGFYPATRSPVSKCTITGWIPAP 300
Db 259 LPSCEKSCDNFYIPNGDYSPLRKHRTGDEITYQCRNGFYPATRGTAKCTSTGWIPAP 318
QY 301 RCLKPCDFPQKHGLYEESSRRFPYFPICKEYSYCDNGFTTTPSOSYWDYLRCTVNG 360
Db 319 RCTLPCDPYDKHGLYHNNRRFPYFPVAVGKYCYCDEHFETPSGYSWDHIHCTQDG 378
QY 361 WEPEVPCRLQCIHFHYVEGESSYWRORYTEGOSAKVQCHSGVSLPNGDITYCTENGWSP 420
Db 379 WSPVPCRLKCYFPLENGYNQNHGKRFVQGSIDVACHPGTALPRAQTVTTCMENGWSP 438
QY 421 PPKCVRIK 428
Db 439 TPRCIRVK 446
```

```
RESULT 3
CFAH_BOVIN STANDARD; PRT; 685 AA.
AC Q28085;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Complement factor H (H factor 1) (Fragments).
GN HF1.
```

Query Match

51.9%; Score 1272.5; DB 1; Length 685;

```
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE OF 17-685 FROM N.A., AND SEQUENCE OF 1-16.
RC TISSUE=Liver;
RX MEDLINE=96202005; PubMed=8615824;
RA Soames C.J., Day A.J., Sim R.B.;
RT "Prediction from sequence comparisons of residues of factor H involved
in the interaction with complement component C3b.";
RL Biochem. J. 315:523-531(1996).
CC C3b by factor I and also increases the rate of dissociation of the
CC C3bB complex (C3 convertase) and the (C3b)NBB complex (C5
CC convertase) in the alternative complement pathway (By similarity).
CC -1- SIMILARITY: CONTAINS AT LEAST 13 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X98697; CAA67257.1; -.
DR HSSP: P10998; 1VVD.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 11.
DR SMART: SM00032; CCP; 11.
KW Complement alternate pathway; Plasma; Repeat; Sushi.
FT NON_TER 1 1
FT NON_CONS 16 17
FT DOMAIN <17 67 SUSHI 2.
FT DOMAIN 70 131 SUSHI 3.
FT DOMAIN 134 188 SUSHI 4.
FT DOMAIN 191 246 SUSHI 5.
FT DOMAIN 249 309 SUSHI 6.
FT DOMAIN 311 366 SUSHI 7.
FT DOMAIN 370 429 SUSHI 8.
FT DOMAIN 431 488 SUSHI 9.
FT DOMAIN 491 547 SUSHI 10.
FT DOMAIN 553 609 SUSHI 11.
FT DOMAIN 614 668 SUSHI 12.
FT DOMAIN 675 >685 SUSHI 13.
FT DISULFID 39 66 BY SIMILARITY.
FT DISULFID 71 117 BY SIMILARITY.
FT DISULFID 103 130 BY SIMILARITY.
FT DISULFID 135 176 BY SIMILARITY.
FT DISULFID 162 187 BY SIMILARITY.
FT DISULFID 192 234 BY SIMILARITY.
FT DISULFID 219 245 BY SIMILARITY.
FT DISULFID 250 297 BY SIMILARITY.
FT DISULFID 280 308 BY SIMILARITY.
FT DISULFID 312 354 BY SIMILARITY.
FT DISULFID 339 365 BY SIMILARITY.
FT DISULFID 371 417 BY SIMILARITY.
FT DISULFID 400 428 BY SIMILARITY.
FT DISULFID 432 476 BY SIMILARITY.
FT DISULFID 459 487 BY SIMILARITY.
FT DISULFID 492 534 BY SIMILARITY.
FT DISULFID 520 546 BY SIMILARITY.
FT DISULFID 554 597 BY SIMILARITY.
FT DISULFID 583 608 BY SIMILARITY.
FT DISULFID 615 656 BY SIMILARITY.
FT DISULFID 642 667 BY SIMILARITY.
FT NON_TER 685 685
SQ SEQUENCE 685 AA; 77536 MW; 69FC9DC8D530E872 CRC64;
```


DR HSSP; P10998; 1VVD.
 DR MIN; 120830; -.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; sushi; 8.
 DR SMART: SM00032; CCP; 8.
 KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal;
 KW Polymorphism.
 FT SIGNAL 1 48
 FT CHAIN 49 597
 FT DOMAIN 49 109
 FT SUSHI 1.
 FT DOMAIN 112 171
 FT SUSHI 2.
 FT DOMAIN 174 235
 FT SUSHI 3.
 FT DOMAIN 238 295
 FT SUSHI 4.
 FT DOMAIN 298 361
 FT SUSHI 5.
 FT DOMAIN 364 423
 FT SUSHI 6.
 FT DOMAIN 425 481
 FT SUSHI 7.
 FT DOMAIN 483 539
 FT SUSHI 8.
 FT BY SIMILARITY.
 FT DISULFID 50 96
 FT DISULFID 81 108
 FT BY SIMILARITY.
 FT DISULFID 113 154
 FT BY SIMILARITY.
 FT DISULFID 140 170
 FT BY SIMILARITY.
 FT DISULFID 175 217
 FT BY SIMILARITY.
 FT DISULFID 203 234
 FT BY SIMILARITY.
 FT DISULFID 239 281
 FT BY SIMILARITY.
 FT DISULFID 267 294
 FT BY SIMILARITY.
 FT DISULFID 299 348
 FT BY SIMILARITY.
 FT DISULFID 332 360
 FT BY SIMILARITY.
 FT DISULFID 364 387
 FT BY SIMILARITY.
 FT DISULFID 365 409
 FT BY SIMILARITY.
 FT DISULFID 422 468
 FT BY SIMILARITY.
 FT DISULFID 454 480
 FT BY SIMILARITY.
 FT DISULFID 484 525
 FT BY SIMILARITY.
 FT DISULFID 511 538
 FT BY SIMILARITY.
 FT DISULFID 546 558
 FT INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
 FT DISULFID 558 588
 FT INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
 FT CARBOHYD 221 221
 FT N-LINKED (GLCNAC. . .).
 FT CARBOHYD 506 506
 FT N-LINKED (GLCNAC. . .).
 FT CARBOHYD 528 528
 FT N-LINKED (GLCNAC. . .).
 FT VARIANT 92 92
 FT Q -> T.
 FT /FTid=VAR_001977.
 FT VARIANT 357 357
 FT Y -> H.
 FT /FTid=VAR_001978.
 FT VARIANT 473 473
 FT W -> L (IN DBSNP:1801341).
 FT /FTid=VAR_012038.
 SQ SEQUENCE 597 AA; 67033 MW; 67E03F2EA85A16DD CRC64;

Query Match 19.68; Score 481.5; DB 1; Length 597;
 Best Local Similarity 26.58; Pred. No. 6.3e-31;
 Matches 134; Conservative 78; Mismatches 195; Indels 99; Gaps 25;

Qy 2 DCKGPPPRENSSEILSGSWSEQLYSEGTOATYKCRPGY-RTLGITIVKVG-KNGEWVPSNPS 59
 Db 49 NCGPPPTLSFAAPMDITLTERFKTGTLLKTLCPGYVRSHSTQTLTNCSDGEWYNT-- 106
 Qy 60 RICKRRCGHPGDPFGSFRVLAVGSEFFEGFAGVYTCDEGYQLGELIDYRECDAD----G 115
 Db 107 -FCIVKCRHPGELRNG--QVEIKTDLFSGQIEFSCSEGFELIGSTTSR-CEVQDRCVG 162
 Qy 116 WTNDIPICEVVKLPVTELENGRIVSGAAEPDQYEFYQGVVFCNCSFKIEGQKEMHCS 175
 Db 163 WSHPLQCEIVKCRPPPPDIRNGR----HSGEENFYAGFVSFYTSCDPRFSLGHASISCT 218
 Qy 176 -EN---GLWSNEKQCQVCEISCLPRVENGDIY-LKPYKENERFQVKCKGQGVYKRGD 230
 Db 219 VENETIGWRSPPTCEKITCRKPDVSHGEMVSGFGPIYKNDIVFKCQKGFVL--RGS 276
 Qy 231 AV--C-TGSGNNO-PSCEENMTCLT-PYIPNGIYT--PHRIK---HRIDDEIRYECKNGF 280
 Db 277 SVIHCDADSKNPNPPACPEPNCINLPDIPHASMETYPRPTKEDVYVVGTVLRYRCHPGY 336
 Qy 281 YPATRSPVSKCTING--WIPAPRCSLAKPCDPQPKHGRLYEESRRP----- 325

Db 337 KPTTDEPTTVICQKNLRWTPYQCEALCCPEPKLNNGEITQHRKSRPANHCVYFYGDEIS 396
 Qy 326 -----YFPVPIG-----KEYSYICDN 341
 Db 397 FSCHEFSRFSALCOGDGTWSPRTPSCGDICNFPPIAHGHYKOSYSYFKEEIIYECOK 456
 Qy 342 GFTTSSQSYWMDYLRCVTNGWEPEVP-CLRQCIFHYVEVEGESSYWMORRYTEGQSAKVQCHS 400
 Db 457 GYILVQGA---KLSCSYSHWSAPAPQCKALCRKPELVNGLRSLVDKQDYVEPVNTVTCQDS 513
 Qy 401 GYSLPNGQDTYYCTEN-GWSPP-PKC 424
 Db 514 GYGVV-GPQSIITCSGNRTWYEPVKC 538
 RESULT 5
 FHR3_HUMAN
 ID FHR3_HUMAN STANDARD; PRT; 330 AA.
 AC Q02985; Q9UJ16;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Complement factor H-related protein 3 precursor (FHR-3) (H factor-like protein 3) (DOWN16).
 DE FHR3.
 GN FHR3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RC TISSUE=Liver;
 RP MEDLINE=93155112; PubMed=8428964;
 RA Skerka C., Kuehn S., Guenther K., Lingelbach K., Zipfel P.F.;
 RT "A novel short consensus repeat-containing molecule is related to human complement factor H";
 RL J. Biol. Chem. 268:2904-2908(1993).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20245597; PubMed=10781834;
 RA Male D.A., Ormsby R.J., Ranganathan S., Giannakis E., Gordon D.L.;
 RT "Complement factor H: sequence analysis of 221 kb of human genomic DNA containing the entire fh, fhr-1 and fhr-3 genes.";
 RL Mol. Immunol. 37:41-52(2000).
 RN (3)
 RP SEQUENCE FROM N.A.
 RA Hall R.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN (4)
 RP REVIEW.
 RX MEDLINE=94226679; PubMed=8172644;
 RA Zipfel P.F., Skerka C.;
 RT "Complement factor H and related proteins: an expanding family of complement-regulatory proteins?";
 RL Immunol. Today 15:121-126(1994).
 CC -!- FUNCTION: MIGHT BE INVOLVED IN COMPLEMENT REGULATION.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
 CC -!- SIMILARITY: STRONG, TO FACTOR H.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X68679; CAA48639.1; -.
 DR EMBL; AL049741; CAB53064.1; -.
 DR PIR; A45222; A45222.
 DR HSSP; P10998; 1VVD.

[illegible]


```

RESULT 8
CRL_HUMAN
ID CRI_HUMAN STANDARD; PRT; 2039 AA.
AC P17927;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement receptor type 1 precursor (C3b/C4b receptor) (CD35
antigen).
GN CRI OR C3BR.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RA MEDLINE=89035992; PubMed=2972794;
RA Klickstein L.B., Bartow T.J., Miletic V., Rabson L.D., Smith J.A.,
RA Fearon D.T.;
RT Identification of distinct C3b and C4b recognition sites in the
RT human C3b/C4b receptor (CRI, CD35) by deletion mutagenesis.;
RL J. Exp. Med. 168:1699-1717(1988).
RN [2]
SEQUENCE OF 503-2039 FROM N.A.
RA MEDLINE=87168191; PubMed=2951479;
RA Klickstein L.B., Wong W.W., Smith J.A., Weis J.H., Wilson J.G.,
RA Fearon D.T.;
RT "Human C3b/C4b receptor (CRI). Demonstration of long homologous
RT repeating domains that are composed of the short consensus repeats
RT characteristics of C3/C4 binding proteins.";
RL J. Exp. Med. 165:1095-1112(1987).
RN [3]
SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.
RA MEDLINE=86067975; PubMed=2933745;
RA Wong W.W., Klickstein L.B., Smith J.A., Weis J.H., Fearon D.T.;
RT Identification of a partial cDNA clone for the human receptor for
RT complement fragments C3b/C4b.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:7711-7715(1985).
CC !- FUNCTION: CRI, PRESENT ON ERYTHROCYTES, LEUKOCYTES, GLOMERULAR
CC PODOCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS, MEDIATES THE
CC BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE
CC ACTIVATED COMPLEMENT.
CC !- SUBUNIT: MONOMER.
CC !- SUBCELLULAR LOCATION: Type 1 membrane protein.
CC !- POLYMORPHISM: CRI IS RESPONSIBLE FOR THE KNOPS BLOOD GROUP SYSTEM.
CC !- MISCELLANEOUS: SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A
CC LONG HOMOLOGOUS REPEAT (LHR). THE N-TERMINAL TWO SRCS OF LHR-A
CC CONTAINED A SITE DETERMINING C4 SPECIFICITY. AND THE N-TERMINAL
CC TWO SRCS OF LHR-B AND -C EACH HAD A SITE DETERMINING C3
CC SPECIFICITY.
CC !- MISCELLANEOUS: THIS IS THE SEQUENCE OF THE F ALLOTYPE OF CRI.
CC !- SIMILARITY: CONTAINS 30 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11569; AA052297.1; -
DR EMBL; M11617; AA052298.1; -
DR EMBL; M11618; AA052299.1; -
DR EMBL; Y00816; CAA68755.1; -
DR EMBL; X05309; CAA28933.1; -
DR PIR; A28507; A28507.
DR PIR; A24748; A24748.
DR PIR; B24748; B24748.
DR PIR; C24748; C24748.
DR PIR; S03843; S03843.
DR HSSP; P08603; 1HFI.
DR MIM; 120620; -

```

```

DR InterPro; IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00084; sush1; 30.
DR SMART; SM00032; CCP; 30.
KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW Receptor; Sush1; Blood group antigen.
FT SIGNAL 1 41
FT CHAIN 42 2039 COMPLEMENT RECEPTOR TYPE 1.
FT DOMAIN 42 1971 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1972 1996 POTENTIAL.
FT DOMAIN 1997 2039 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 42 42 PYROLIDONE CARBOXYLIC ACID (POTENTIAL).
FT DOMAIN 42 100 SUSHI A1.
FT DOMAIN 103 162 SUSHI A2.
FT DOMAIN 165 233 SUSHI A3.
FT DOMAIN 237 294 SUSHI A4.
FT DOMAIN 296 354 SUSHI A5.
FT DOMAIN 357 417 SUSHI A6.
FT DOMAIN 420 488 SUSHI A7.
FT DOMAIN 492 550 SUSHI B1.
FT DOMAIN 553 612 SUSHI B2.
FT DOMAIN 615 683 SUSHI B3.
FT DOMAIN 687 744 SUSHI B4.
FT DOMAIN 746 804 SUSHI B5.
FT DOMAIN 807 867 SUSHI B6.
FT DOMAIN 870 938 SUSHI B7.
FT DOMAIN 942 1000 SUSHI C1.
FT DOMAIN 1003 1062 SUSHI C2.
FT DOMAIN 1065 1133 SUSHI C3.
FT DOMAIN 1137 1194 SUSHI C4.
FT DOMAIN 1196 1254 SUSHI C5.
FT DOMAIN 1257 1317 SUSHI C6.
FT DOMAIN 1320 1388 SUSHI C7.
FT DOMAIN 1395 1453 SUSHI D1.
FT DOMAIN 1456 1515 SUSHI D2.
FT DOMAIN 1518 1585 SUSHI D3.
FT DOMAIN 1590 1647 SUSHI D4.
FT DOMAIN 1649 1707 SUSHI D5.
FT DOMAIN 1710 1770 SUSHI D6.
FT DOMAIN 1773 1841 SUSHI D7.
FT DOMAIN 1847 1905 SUSHI E1.
FT DOMAIN 1908 1966 SUSHI E2.
FT DISULFID 43 86 BY SIMILARITY.
FT DISULFID 73 99 BY SIMILARITY.
FT DISULFID 104 145 BY SIMILARITY.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 195 232 BY SIMILARITY.
FT DISULFID 238 280 BY SIMILARITY.
FT DISULFID 266 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 358 400 BY SIMILARITY.
FT DISULFID 386 416 BY SIMILARITY.
FT DISULFID 421 470 BY SIMILARITY.
FT DISULFID 450 487 BY SIMILARITY.
FT DISULFID 493 536 BY SIMILARITY.
FT DISULFID 523 549 BY SIMILARITY.
FT DISULFID 554 595 BY SIMILARITY.
FT DISULFID 581 611 BY SIMILARITY.
FT DISULFID 616 665 BY SIMILARITY.
FT DISULFID 645 682 BY SIMILARITY.
FT DISULFID 688 730 BY SIMILARITY.
FT DISULFID 716 743 BY SIMILARITY.
FT DISULFID 747 790 BY SIMILARITY.
FT DISULFID 776 803 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 866 BY SIMILARITY.
FT DISULFID 871 920 BY SIMILARITY.
FT DISULFID 900 937 BY SIMILARITY.
FT DISULFID 943 986 BY SIMILARITY.
FT DISULFID 973 999 BY SIMILARITY.
FT DISULFID 1004 1045 BY SIMILARITY.
FT DISULFID 1031 1061 BY SIMILARITY.

```

```
FT DISULFID 1066 BY SIMILARITY. 1115
FT DISULFID 1095 BY SIMILARITY. 1132
FT DISULFID 1138 BY SIMILARITY. 1180
FT DISULFID 1166 BY SIMILARITY. 1193
FT DISULFID 1197 BY SIMILARITY. 1240
FT DISULFID 1226 BY SIMILARITY. 1253
FT DISULFID 1258 BY SIMILARITY. 1300
FT DISULFID 1286 BY SIMILARITY. 1316
FT DISULFID 1321 BY SIMILARITY. 1370
FT DISULFID 1350 BY SIMILARITY. 1387
FT DISULFID 1396 BY SIMILARITY. 1439
FT DISULFID 1426 BY SIMILARITY. 1452
FT DISULFID 1457 BY SIMILARITY. 1496
FT DISULFID 1484 BY SIMILARITY. 1514
FT DISULFID 1519 BY SIMILARITY. 1568
FT DISULFID 1548 BY SIMILARITY. 1585
FT DISULFID 1591 BY SIMILARITY. 1633
FT DISULFID 1619 BY SIMILARITY. 1646
FT DISULFID 1650 BY SIMILARITY. 1693
FT DISULFID 1679 BY SIMILARITY. 1706
FT DISULFID 1711 BY SIMILARITY. 1753
FT DISULFID 1739 BY SIMILARITY. 1769
FT DISULFID 1774 BY SIMILARITY. 1823
FT DISULFID 1803 BY SIMILARITY. 1840
FT DISULFID 1848 BY SIMILARITY. 1891
FT DISULFID 1877 BY SIMILARITY. 1904
FT DISULFID 1909 BY SIMILARITY. 1952
FT DISULFID 1938 BY SIMILARITY. 1965
FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 509 509 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 860 860 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 959 959 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1028 1028 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1152 1152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1310 1310 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1481 1481 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1504 1504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1540 1540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1605 1605 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1763 1763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1908 1908 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2039 AA; 223589 MW; B2FD29B6AD3C5EB7 CRC64;

Query Match 17.8%; Score 437; DB 1; Length 2039;
Best Local Similarity 27.7%; Pred. No. 8.5e-27;
Matches 132; Conservative 63; Mismatches 149; Indels 132; Gaps 31;

QY 27 GTOATYCRPGVYRTLTITVVKVNGEMVSPNSRICRKPCHGPTPGSGFRVLAVGSEF 86
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   66 GYLVNTECRPGVGRPFISICKLNSVWTGAKDR--CRRKSCRNPDPVNGVHVIRG--I 121
   :|||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 122 QFGSIKYSCTGYRLIGS--SSATCIISGDTYVNDNETICDRICPLPT-ITNGDFIS 179
QY 142 GAEPDQYFGVAVFENSG-----FKIEGQKEMHCSEN-----GLWSNEKPGC-VRI 190
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 180 TNRE---NPHYGSVTVYRCNPGSGGRKVFELVGEPSIYCTSDQVGIWSGPAQCIIIN 236
QY 191 SCLPPRVENG----DGIYLLKPYVKENERFYKCKQKGFVYKRGDAVCTG--SGWNPD--PSC 244
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 237 KCTPPNVENGILVSN---RSLSFLNEVVEFRCPQFGFMKPPRVKRCQALNKWEPLPSC 293
QY 245 EEMTCLTYPYIPNGIYTPHRIKHRIID----DEIRYECKNGFYPATRSPVSKCITG-WIP- 298
   : | | | | : : : : : : : : : : : : : : : : : : : : : : : : | | |
```

```
Db 294 SRV-CQPP--PDVLHAERTQDKDNFSPGOEVFYSCERG-YDLRGAASMRCTPOGDWSPA 349
QY 299 APRCSLAPCD--FPQFKHGRLYEESRRPYFPV--PIGKEYSYCDNGFTT--PSQSYWD 352
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 350 APTCEVAKSCDDFMGQLLNGRV-----LFPVNLQLGAKVDFVCDEGFKLGSSASY-- 399
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 353 YLRCTVNG---WEPEVPCLRQCIFHYVEYGESSYMQRRYIEGSAKVQCHSGYSLPN-- 406
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 400 ---CVLAGMESLWNSVVPCEQ-----IFCPSPPVIPNGR 431
QY 407 -----GQDTYYCTE-----NG-W-SPPKPC 424
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 432 HTGKPLEVFPFGKAVNYTCDPHPDRGT'SFDLIGESTIRCTSDPQCGNGVWSSPAPRC 487
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
C4BP_MOUSE STANDARD: PRT; 469 AA.
ID C4BP_MOUSE AC P08607;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C4b-binding protein precursor (C4bp).
GN C4BPA OR C4BP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=88024937; PubMed=3663616;
RA Kristensen T., Ogata R.T., Chung L.P., Reid K.B.M., Tack B.F.;
RT "cDNA structure of murine C4b-binding protein, a regulatory component
  of the serum complement system.";
RL Biochemistry 26:4668-4674(1987).
CC -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
  ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
  (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
  ALSO ACCELERATES THE DEGRADATION OF THE C4B/C2A COMPLEX (C3
  CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
  CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
CC -!- SUBUNIT: MONOMER; NOT COVALENTLY LINKED. MOUSE LACKS THE
  BETA CHAIN OF C4BP.
CC -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -!- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-44 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC EMBL; M17122; AAA37312.1; ALT_INIT.
DR PIR; A27117; NBMSC4.
DR HSP; P10998; 1VVD.
DR MGD; MGI:88229; C4bp.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 6.
DR SMART; SM00032; CCP; 6.
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT SIGNAL 1 56
FT CHAIN 57 469 C4B-BINDING PROTEIN.
FT DOMAIN 57 116 SUSHI 1.
FT DOMAIN 119 177 SUSHI 2.
FT DOMAIN 180 241 SUSHI 3.
FT DOMAIN 244 300 SUSHI 4.
FT DOMAIN 302 356 SUSHI 5.
FT DOMAIN 358 414 SUSHI 6.
FT DISULFID 58 103 BY SIMILARITY.
FT DISULFID 88 115 BY SIMILARITY.
```

```
FT DISULFID 120 160 BY SIMILARITY.
FT DISULFID 146 176 BY SIMILARITY.
FT DISULFID 181 223 BY SIMILARITY.
FT DISULFID 209 240 BY SIMILARITY.
FT DISULFID 245 287 BY SIMILARITY.
FT DISULFID 273 299 BY SIMILARITY.
FT DISULFID 303 343 BY SIMILARITY.
FT DISULFID 329 355 BY SIMILARITY.
FT DISULFID 359 400 BY SIMILARITY.
FT DISULFID 386 413 BY SIMILARITY.
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 428 428 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 469 AA: 51551 MW: 41E137CB8D8C6321 CRG64;

Query Match 15.7%; Score 384.5; DB 1: Length 469;
Best Local Similarity 29.0%; Pred. No. 2.4e-23;
Matches 112; Conservative 64; Mismatches 161; Indels 49; Gaps 21;

QY 5 GPPPRENSILSGWSEQLYSECTQATYKRCPGY-RTLGTIVKVK-NGEWPSPNPSRIC 62
Db 59 GPPPAIPNALPASDVNRTDFESHITLKYELCPYGRGISMVMVYCKPSEW---EISVSC 115
QY 63 RRRPCGHGDTFPGSLAVGEFEFGAKVVTYCDGYOLLGEIDYRECDAG---WTN 118
Db 116 AKKHCNPGYLDNGYVN---GETITGSIQIEFSCQEGFILVGS-STSSCEVRKGVAMSN 171
QY 119 DIPICEVVKLPTELENGRIVSGAEPDQYVFGQVRFECNSGRKIEGQKEMHCSNG 178
Db 172 PPPECVIVKCGPPDISNGK-HSGT---EDFPYNNHGISITCDPGRPLVGSPPFIGTVVN 227
QY 179 ----LMSNEKPCQVEISCLPPRVENGDIY-LKPVYKENERFOYKCKGCFVYKERGDAVC 233
Db 228 KIVPWSSSPPTCEKICISQNLHGVIVSGYKATYTHRDSVRLACLNGTLVLRIVIEC 287
QY 234 TSGG-WNPQSPCEMTLPYIPNGIYTPHRIKHRIDDEIRYCKNGFYPATRSPYSKCT 292
Db 288 QGNGNSSLPTCEFCDLPPAIVNGYIT-SMYSKI-TLVTVCECDKGYRLVCKAIIIS-CS 344
QY 293 ITGWI-PAPCSLKCDPQFQKHGL-----YEEERRPYFPVPIGKEYSYCDNGFT-T 345
Db 345 FSKWGTAPCKAL-CQKPEVNGTLSDKQDYVES-----ENVTIQDSGFAML 393
QY 346 PSQSYWDYLRCVNG-WEPEVPCLRQ 370
Db 394 GSQS-----ISCSSESGTWYEVPRCEQ 415
```

```
RESULT 10
CR2_MOUSE STANDARD; PRT: 1025 AA.
AC CR2_MOUSE
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Complement receptor type 2 precursor (Cr2) (Complement C3d receptor).
GN CR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C.
RX MEDLINE-90229735; Pubmed-2139457;
RA Fingerth J.D.;
RT "Comparative structure and evolution of murine CR2. The homolog of
the human C3d/EBV receptor (CD21).";
```

```
RL J. Immunol. 144:3458-3467(1990).
RN [2]
RP SEQUENCE OF 12-1025 FROM N.A.
RX MEDLINE-91010789; Pubmed-2145366;
RA Molina H., Kinoshita T., Inoue K., Carel J.C., Holers V.M.;
RT "A molecular and immunochemical characterization of mouse CR2.
Evidence for a single gene model of mouse complement receptors 1 and
2.";
RN J. Immunol. 145:2974-2983(1990).
RN [3]
RP SEQUENCE OF 343-401 AND 991-1025 FROM N.A.
RX MEDLINE-89098890; Pubmed-2783485;
RA Fingerth J.D., Benedict M.A., Levy D.N., Strominger J.L.;
RT "Identification of murine complement receptor type 2.";
Proc. Natl. Acad. Sci. U.S.A. 86:242-246(1989).
RN [4]
RP SEQUENCE OF 289-1025 FROM N.A.
RX MEDLINE-89381350; Pubmed-2528587;
RA Kurtz C.B., Paul M.S., Regerter M., Weis J.J., Weis J.H.;
RT "Murine complement receptor gene family. II. Identification and
characterization of the murine homolog (Cr2) to human CR2 and its
molecular linkage to Cr2.";
J. Immunol. 143:2058-2067(1989).
CC - FUNCTION: RECEPTOR FOR COMPLEMENT C3D. PARTICIPATES IN B
LYMPHOCYTES ACTIVATION.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - TISSUE SPECIFICITY: B LYMPHOCYTES.
CC - SIMILARITY: TO HUMAN C3D/EBV RECEPTOR (CD21).
CC - SIMILARITY: CONTAINS 15 SUSHI (SCR) DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; M81083; AAA374451.1; -
EMBL; M35684; AAA37448.1; -
EMBL; M61132; AAA63295.1; -
EMBL; M35685; AAA37450.1; ALT_SEQ.
EMBL; M29281; AAA37447.1; -
PIR; A43526; A43526.
HSP; P10998; LVND.
MGD; MGI:88489; Cr2.
InterPro; IPR000436; Sush1_SCR_CCP.
Pfam; PF00084; sush1; 14.
SMART; SM00032; CCP; 14.
Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
Receptor; Sush1.
SIGNAL 1 11
CHAIN 12 1025 COMPLEMENT RECEPTOR TYPE 2.
DOMAIN 12 963 EXTRACELLULAR (POTENTIAL).
TRANSMEM 964 990 POTENTIAL.
DOMAIN 991 1025 CYTOPLASMIC (POTENTIAL).
DOMAIN 13 74 SUSHI 1.
DOMAIN 81 139 SUSHI 2.
DOMAIN 145 203 SUSHI 3.
DOMAIN 206 264 SUSHI 4.
DOMAIN 267 335 SUSHI 5.
DOMAIN 342 399 SUSHI 6.
DOMAIN 401 458 SUSHI 7.
DOMAIN 462 515 SUSHI 8.
DOMAIN 518 586 SUSHI 9.
DOMAIN 593 649 SUSHI 10.
DOMAIN 653 705 SUSHI 11.
DOMAIN 708 769 SUSHI 12.
DOMAIN 777 834 SUSHI 13.
DOMAIN 840 898 SUSHI 14.
DOMAIN 901 959 SUSHI 15.
DISULFID 14 56 BY SIMILARITY.
DISULFID 42 73 BY SIMILARITY.
```

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; M26004; AAB35786.1; -
 DR EMBL; M26016; AAB04638.1; -
 DR EMBL; M24007; AAB04638.1; JOINED.
 DR EMBL; M24008; AAB04638.1; JOINED.
 DR EMBL; M24009; AAB04638.1; JOINED.
 DR EMBL; M24010; AAB04638.1; JOINED.
 DR EMBL; M24011; AAB04638.1; JOINED.
 DR EMBL; M26009; AAB04638.1; JOINED.
 DR EMBL; M26010; AAB04638.1; JOINED.
 DR EMBL; M26011; AAB04638.1; JOINED.
 DR EMBL; M26012; AAB04638.1; JOINED.
 DR EMBL; M26013; AAB04638.1; JOINED.
 DR EMBL; M26014; AAB04638.1; JOINED.
 DR EMBL; M26015; AAB04638.1; JOINED.
 DR EMBL; S62696; AAB27186.1; -
 DR PIR; A32036; A32036.
 DR PIR; A24319; A24319.
 DR PIR; B24319; B24319.
 DR PIR; C24319; C24319.
 DR PIR; D24319; D24319.
 DR PIR; E24319; E24319.
 DR HSSP; P10998; 1VVD.
 DR MIN; 120650; -
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00084; sushi; 15.
 DR SMART; SM00032; CCP; 14.
 KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
 KW Receptor; Sushi; Alternative splicing.
 FT CHAIN 1 20
 FT DOMAIN 21 1033
 FT TRANSMEM 21 971
 FT DOMAIN 972 999
 FT DOMAIN 1000 1033
 FT DOMAIN 22 83
 FT DOMAIN 90 147
 FT DOMAIN 153 211
 FT DOMAIN 214 272
 FT DOMAIN 275 343
 FT DOMAIN 350 407
 FT DOMAIN 409 467
 FT DOMAIN 470 523
 FT DOMAIN 526 594
 FT DOMAIN 601 658
 FT DOMAIN 661 715
 FT DOMAIN 718 780
 FT DOMAIN 787 844
 FT DOMAIN 850 908
 FT DOMAIN 911 969
 FT DISULFID 23 65
 FT DISULFID 51 82
 FT DISULFID 91 132
 FT DISULFID 118 146
 FT DISULFID 154 197
 FT DISULFID 183 210
 FT DISULFID 215 256
 FT DISULFID 242 271
 FT DISULFID 276 325
 FT DISULFID 305 342
 FT DISULFID 351 393
 FT DISULFID 379 406
 FT DISULFID 410 453
 FT DISULFID 439 466
 FT DISULFID 471 509

FT DISULFID 495 522 BY SIMILARITY.
 FT DISULFID 527 576 BY SIMILARITY.
 FT DISULFID 556 593 BY SIMILARITY.
 FT DISULFID 602 644 BY SIMILARITY.
 FT DISULFID 657 699 BY SIMILARITY.
 FT DISULFID 662 714 BY SIMILARITY.
 FT DISULFID 685 762 BY SIMILARITY.
 FT DISULFID 719 779 BY SIMILARITY.
 FT DISULFID 748 830 BY SIMILARITY.
 FT DISULFID 788 843 BY SIMILARITY.
 FT DISULFID 816 894 BY SIMILARITY.
 FT DISULFID 851 907 BY SIMILARITY.
 FT DISULFID 912 955 BY SIMILARITY.
 FT DISULFID 941 968 BY SIMILARITY.
 FT CARBOHYD 121 121 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 294 294 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 492 492 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 682 682 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 800 800 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 823 823 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 861 861 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 911 911 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPLIC 499 524 MISSING (IN SHORT ISOFORM).
 FT VARSPLIC 525 556 ITCPPPPVIYNGAHTGSSLEDPYGTVTYTC -> NHLPT
 TPCYLOWGTHREFLRRFSIMNHGLHM (IN SHORT
 ISOFORM).
 FT CONFLICT 567 667 Q -> D (IN REF. 2).
 FT CONFLICT 902 902 Q -> G (IN REF. 2).
 FT CONFLICT 906 906 H -> L (IN REF. 2).
 SQ SEQUENCE 1033 AA; 112973 MW; 1749DBA07847ADA CRC64;
 Query Match 15.0%; Score 368; DB 1; Length 1033;
 Best Local Similarity 24.3%; Pred. No. 1.2e-21;
 Matches 131; Conservative 63; Mismatches 198; Indels 146; Gaps 31;
 QY 1 EDCKGPPPPRENSSEILSGSWSEQ---LYSEGQTATYKCRPGYRTLG-TIVKCKNGSWPS 56
 DB 408 KECQAPP-----NILNGQKEDRHMYRFDGTSIKSCNPGYVYLVGEESIQCTSEGWTTP 462
 QY 57 NPSRCRRKPCGHPGDTFFGSRFVAGSEFEF-GAKVYVTCDEGYOLGEIDYRECDAD- 114
 DB 463 VPQ--CKVAACEATGR-----QLLTQPHQVRPDVANSCCGEGYKLSGV-YQECOGTI 513
 QY 115 GWTNDIPICEVVKCLPVTELENGRIVSGAAEPDQYEGYGVVFECSNG-----FKIEG 168
 DB 514 PFWFIRLCKEITCPPPPVIYNGAHTGSSLE---DFPYGTVTYTCNPGPERGVFSLIG 570
 QY 169 QKEMHCSEN---GLWSNEKPCQ---VEISCLPPRVENGDIYLPV-YKENERFQYKC 219
 DB 571 ESTIRCTSDQERGTSWGPAPLCKLSLLAVQCSHVHIANGYKISKEAPYFYNDVTFKC 630
 QY 220 KQGFYKRGDAVCTGSG-WNPQ-PSCEMTG----- 249
 DB 631 YSGFTLAGSSQIRCKADNTWDPEIPVCEKETCQHVRSQRLQELPAGSRVELVNTSCDQGY 690
 QY 250 LT----- 750
 DB 691 LTGHAYQMCQDAENGWFKKIPCKVICHHPPPVIVNKGHTGMAENFLYGNVESTEDCQ 750
 QY 279 GFYPATRSVPVKCTIT-----GWI-PAPRCSLKP---CDFPQFQKHGRLYXEESRRPYFP 328
 DB 751 GFYLLGKKL-QCRSDSKGHGWSGSPQCLRSPPVTRCPNPEVKG----- 796
 QY 329 VPICKESYI-----CDNGETTPSQSYWDYLRCTVNGWBEVP-CLRQCIFHYVE 377
 DB 797 YKLNKTHSAYSHNDIVVYDCNPGFTMGSR---VIRCHTDNTWVPVPTCKKAGICPP 853
 QY 378 YGESSYNOR-----RYIEGQSAKQVCSYSLNPGQDTYYCTENG-WS-PPKCVRI 427

Db 425 C-NEYLLRGSKISRCBOGKSSPPVC-LEPCTVNVDMNRRNEMKWKYCKVLHGDLI 482
QY 336 SYCNDGFTTPSQSYDWYLCRTVNGWEPEVP-CLRQ-----CIFYVEYGESSYW 384
Db 483 DFVQKQGYDLSPTLSELSEVOCNGEVKYPCLTRKESKGMCTSPPLIKHGVIISST--- 539
QY 385 QRRYIEGOSAKVOCHGSLPQNGQDTYYCTENGWSPPKCV 425
Db 540 VDTYENGSSVEYRCFDDHFFLEGSREA-YCLDGMWTTPLCL 579

RESULT 13
FHR4_HUMAN
ID FHR4_HUMAN STANDARD: PRT: 331 AA.
AC Q92496; Q90JY6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement factor H-related protein 4 precursor (FHR-4).
GN FHR4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97190290; PubMed=9038172;
RA Skerka C., Hellwege J., Weber W., Tilkorn A., Buck F., Marti T.,
RA Kampen E., Beisiegel U., Zipfel P.F.;
RT "The human factor H-related protein 4 (FHR-4). A novel short consensus
RT repeat-containing protein is associated with human triglyceride-rich
RT lipoproteins.";
RL J. Biol. Chem. 272:5627-5634(1997).
RN [2]
RP SEQUENCE OF 1-19 FROM N.A.
RA Male D.A., Ormsby R.J., Giannakis E., Gordon D.L.;
RT "Promoter region of complement factor H-related 4 (fhr-4) gene.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RX CHARACTERIZATION.
RP MEDLINE=98136603; PubMed=9476126;
RA Hellwege J., Skerka C., Zipfel P.F.;
RT "Biochemical and functional characterization of the factor-H-related
RT protein 4 (FHR-4).";
RL Immunopharmacology 38:149-157(1997).
CC -!- FUNCTION: INVOLVED IN COMPLEMENT REGULATION. CAN ASSOCIATE WITH
CC LIPOPROTEINS AND MAY PLAY A ROLE IN LIPID METABOLISM.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- PTM: GLYCOSYLATED.
CC -!- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC -!- SIMILARITY: STRONG, TO FACTOR H.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X98337; CAA66980.1; ..
CC EMBL; AF190816; AAF05951.1; ..
CC HSSP; P10998; 1VVD.
CC MIN; 605337; ..
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00084; sushi; 4.
CC SMART; SM00032; CCP; 4.
KW Repeat; Glycoprotein; Sushi; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 331 COMPLEMENT FACTOR H-RELATED PROTEIN 4.
FT DOMAIN 20 84 SUSHI 1.

FT DOMAIN 95 146 SUSHI 2.
FT DOMAIN 147 205 SUSHI 3.
FT DOMAIN 206 266 SUSHI 4.
FT DOMAIN 267 331 SUSHI 5.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 331 AA; 37325 MW; 5A0D04AB4B841424 CRC64;

Query Match 13.5%; Score 332.5; DB 1; Length 331;
Best Local Similarity 44.6%; Pred. No. 2.2e-19;
Matches 62; Conservative 19; Mismatches 53; Indels 5; Gaps 3;

QY 293 ITGWIP-APRCSLKPCDFPQFKHRLYYEESRRPVPVPIGKEYSYICNGFTTPSQSYW 351
Db 9 LTLWVSCANGQEVKPCDFEIQHGLYKSLRLYFPAAAGQSYICYCDQNFVTPSGSYW 68
QY 352 DYLCRTVNGWEPEVPCRLQCIFH--YVEYGESSYWMORRYIEGOSAKVCHGSLPQNGD 409
Db 69 DYIHCTQDGLVTPCLRTCSKSDIEIENGFISSSSIVILNKIEIQYCKPKGYATADGNS 128
QY 410 --YYCYCTENGWSPPKCV 426
Db 129 SCSITCLQNGWSAQPTCIK 147

RESULT 14
F13B_MOUSE
ID F13B_MOUSE STANDARD: PRT: 668 AA.
AC Q07968;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coagulation factor XIII B chain precursor (EC 2.3.2.13) (Protein-
DE glutamine gamma-glutamyltransferase B chain) (Transglutaminase B
DE chain).
GN F13B OR CF13B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B10.D2/OSN; TISSUE=Liver;
RX MEDLINE=93224141; PubMed=8468048;
RA Nonaka M., Matsuda Y., Shiroishi T., Moriaki K., Nonaka M.,
RA Natsuume-Sakai S.;
RT "Molecular cloning of the b subunit of mouse coagulation factor XIII
RT and assignment of the gene to chromosome 1: close evolutionary
RT relationship to complement factor H.";
RL Genomics 15:535-542(1993).
CC -!- FUNCTION: THE B CHAIN OF FACTOR XIII IS NOT CATALYTICALLY ACTIVE,
CC BUT IS THOUGHT TO STABILIZE THE A SUBUNIT AND REGULATE THE RATE
CC OF TRANSGLUTAMINASE FORMATION BY THROMBIN.
CC -!- SUBUNIT: TETRAMER OF TWO A CHAINS AND TWO B CHAINS.
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LIVER AND KIDNEY.
CC -!- SIMILARITY: CONTAINS 10 SUSHI (SCR) DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; D10071; BAA00963.1; ..
CC PIR; A46013; A46013.
CC HSSP; P08603; 1HFI.
CC MGI; MGI:88379; F13b.
CC InterPro; IPR000436; Sushi_SCR_CCP.

DR Pfam: PF00084; sush1; 8.
DR SMART; SM0032; CCP; 8.
KW Transferase; Plasma; Blood coagulation; Repeat; Glycoprotein; Signal;
KW Sush1.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 668
FT DOMAIN 24 88
FT DOMAIN 90 147
FT DOMAIN 152 209
FT DOMAIN 212 268
FT DOMAIN 273 328
FT DOMAIN 335 390
FT DOMAIN 395 451
FT DOMAIN 453 516
FT DOMAIN 523 579
FT DOMAIN 581 647
FT DISULFID 25 76
FT DISULFID 59 87
FT DISULFID 91 135
FT DISULFID 118 146
FT DISULFID 153 197
FT DISULFID 180 208
FT DISULFID 213 255
FT DISULFID 241 267
FT DISULFID 302 327
FT DISULFID 336 378
FT DISULFID 364 389
FT DISULFID 396 439
FT DISULFID 425 450
FT DISULFID 454 505
FT DISULFID 486 515
FT DISULFID 524 567
FT DISULFID 553 578
FT DISULFID 582 636
FT DISULFID 616 646
FT CARBOHYD 162 162
FT CARBOHYD 545 545
FT SEQUENCE 668 AA; 76078 MW; 80BC9E00A9E53FA6 CRC64;
Query Match 13.5%; Score 332; DB 1; Length 668;
Best Local Similarity 27.0%; Pred. No. 5,3e-19;
Matches 129; Conservative 48; Mismatches 199; Indels 102; Gaps 29;
QY 1 EDCGPPPRENSELGWSQLYSEGTOATYK-----CRPYRTL---GTIVKVKC 49
DB 151 ETCLAP-----ELEHGNYST-----TORTEKVKDIAVYCTAGYTTTGKOTGEAECQ 198
QY 50 NGEW--VPSNPSRICRKRCPGHPGDPFGSFR-----AVGSEFEFGAKVYVTCDEGY 100
DB 199 ANGWSLTQ-----CNKLMC-----SSLRIENGYPHPVKQTYEGDLVQFFCHENY 245
QY 101 QLLGEIDYRECDADGWTNDIPICE--VVKCLPVTLENGRIVSGAAEP--DOEYVFGOVVR 157
DB 246 YLSSG--DLIOCFYNGFSPESPICEGRNRNCRPPPPVPLNSKI-----QPHSTTYRGERVH 299
QY 158 FPCNSGFIEGQKHEKHCSENGLSNEKQOCVE-----ISC-LPPRVENGDIYKLPVYKEN 212
DB 300 IECENLFVIQSGEELLC-ENGKWT-EPPKCIIEKEKVAECQPPSVENGVAHPHSEIYISG 357
QY 213 EFQYKCKQGFYKRGDAVCTGSGWNPQPCSE--MTCL-TPYIPNGIYTPHRI-KHRI 268
DB 358 DKVYRCGGYSLRGSSTITCNRGWTLPPECVENIENCKPPDPDIANGVVVDGLLASYTT 417
QY 269 DDEIYEKNGFYPATRSPVSKCTITGWIPAPRCSLKPCDFP-----QFK----- 313
DB 418 GSSVEYRC-NEYLLKGSSETRCEQAGWSSPPVC-LEPCTIDVDHMRNRNIOLKWKYEK 475
QY 314 --HGLRYEESRRPYF---PVPICKEYSYCDNGTFTSQSWYDLVLRCTVNGWEPEVPC 368
DB 476 ILHGLDIDVCKQGNLSPSIPL-SEISAQNRG---DVRYPMCIRKESKGMCAAPV 530

QY 369 ROCIFHYVEYGE--SSYQWRRYIEGQSAKVOCHSGYSLPQNGDYYCTENGWSPPPKCV 425
DB 531 RN-----GDIVSSAARTYENGSSVEYRCFDNHEL-QGSONVYCVGVWTTTPPSCL 579

RESULT 15
FHRL_HUMAN
ID FHRL_HUMAN STANDARD; PRT; 330 AA.
AC Q03591; Q0J17;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement factor H-related protein 1 precursor (FHR-1) (H factor-like protein 1) (H-factor like 1) (H36).
OS Homo sapiens (Human).
GN FHR1 OR HFL1 OR CFHL.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
TX TISSUE=Liver;
RX MEDLINE=91201892; PubMed=1826708;
RA Estaller C., Koistinen V., Schwaeble W., Dierich M.P., Weiss E.H.;
RT "Cloning of the 1.4-kb mRNA species of human complement factor H reveals a novel member of the short consensus repeat family related to the carboxy terminal of the classical 150-kDa molecule.";
RL J. Immunol. 146:3190-3196(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20245597; PubMed=10781834;
RA Male D.A., Ormsby R.J., Ranganathan S., Giannakis E., Gordon D.L.;
RT "Complement factor H: sequence analysis of 221 kb of human genomic DNA containing the entire fh, FHR-1 and FHR-3 genes.";
RL Mol. Immunol. 37:41-52(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Hall R.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 4-330 FROM N.A.
RX MEDLINE=91268081; PubMed=1711047;
RA Skerka C., Horstmann R.D., Zipfel P.F.;
RT "Molecular cloning of a human serum protein structurally related to complement factor H.";
RL J. Biol. Chem. 266:12015-12020(1991).
RN [5]
RP CARBOHYDRATES.
RX MEDLINE=91123700; PubMed=1825108;
RA Timmann C., Leippe M., Horstmann R.D.;
RT "Two major serum components antigenically related to complement factor H are different glycosylation forms of a single protein with no factor H-like complement regulatory functions.";
RL J. Immunol. 146:1265-1270(1991).
RN [6]
RP REVIEW.
RX MEDLINE=94226679; PubMed=8172644;
RA Zipfel P.F., Skerka C.;
RT "Complement factor H and related proteins: an expanding family of complement-regulatory proteins?";
RL Immunol. Today 15:121-126(1994).
CC -!- FUNCTION: MIGHT BE INVOLVED IN COMPLEMENT REGULATION. CAN ASSOCIATE WITH LIPOPROTEINS AND MAY PLAY A ROLE IN LIPID METABOLISM.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: LIVER.
CC -!- PTM: N-GLYCOSYLATED. TWO FORMS ARE OBSERVED; ONE WITH A SINGLE SIDE-CHAIN AND THE OTHER WITH TWO.
CC -!- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC -!- SIMILARITY: STRONG, TO FACTOR H.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:11:41 ; Search time 67.22 Seconds
(without alignments)
1101.485 Million cell updates/sec

Title: US-09-316-163-14

Perfect score: 2454

Sequence: 1 EDCKGPPPRENSEILLGSWS.....DTYYCTENGWSPPKCVRIK 428

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2426	98.9	1236	11 Q91YB6	Q91YB6 rattus norv
2	1714	69.8	1172	4 Q9NU87	Q9NU87 homo sapien
3	1242	50.6	669	6 Q28085	Q28085 bos taurus
4	813	33.1	452	11 Q61407	Q61407 mus musculus
5	811	33.0	303	11 Q61405	Q61405 mus musculus
6	569.5	23.2	1053	13 Q91275	Q91275 paralabrax
7	566	23.1	808	11 Q61408	Q61408 mus musculus
8	478	19.5	555	11 Q99JAI	Q99JAI cavia porce
9	466	19.0	559	11 Q63135	Q63135 rattus norv
10	453	18.5	679	11 Q99254	Q99254 mus musculus
11	447.5	18.2	522	6 Q28769	Q28769 papio cynoc
12	447	18.2	2014	6 Q29530	Q29530 pan troglod
13	437	17.8	559	4 Q9UGV2	Q9UGV2 homo sapien
14	437	17.8	2039	4 Q16745	Q16745 homo sapien
15	437	17.8	2489	4 Q16744	Q16744 homo sapien
16	436	17.8	1911	6 Q29528	Q29528 papio hamad

ALIGNMENTS

RESULT 1

Q91YB6 ID Q91YB6 PRELIMINARY; PRT; 1236 AA.
AC Q91YB6
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
D1 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE COMPLEMENT INHIBITORY FACTOR H.
GN FH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RA Demberg T., Goetze O., Schlaf G.;
RT "Rat complement factor H: molecular cloning, sequencing and expression in tissues and isolated cells."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ320522; CAC67513.1; -
SQ SEQUENCE 1236 AA; 140343 MW; 1AC89FFA28232EBF CRC64;

Query Match 98.9%; Score 2426; DB 11; Length 1236;
Best Local Similarity 99.1%; Pred. No. 3.9e-215;
Matches 424; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 EDCKGPPPRENSEILLGSWSSEQLYSEGTQATYKCRPGVYRTLTGTIVKVKNGEWVPSNPSR 60
|||||
Db 19 EDCKGPPPRENSEILLGSWSSEQLYSEGTQATYKCRPGVYRTLTGTIVKVKNGEWVPSNPSR 78
|||||
QY 61 ICRKRCGHPGDPFPFGSFLAVGSEFEFGARVYVTCDEGYQLLGEIDYRECDADGWTNDI 120
|||||
Db 79 ICRKRCGHPGDPFPFGSFLAVGSEFEFGARVYVTCDEGYQLLGEIDYRECDADGWTNDI 138
|||||
QY 121 PICEVVKCLPVTLENGRIVSGAAEPDQYFYFGVVFECNSGFKIEGQKMHCSENGLW 180
|||||
Db 139 PICEVVKCLPVTLENGRIVSGAAEPDQYFYFGVVFECNSGFKIEGQKMHCSENGLW 198

```
Qy 181 SNEKPCQVEISCLPPRVENGDIYLLKPVYKENERFOYKCKOGFYKRGDAVCTGSGWNP 240
Db 199 SNEKPCQVEISCLPPRVENGDIYLLKPVYKENERFOYKCKOGFYKRGDAVCTGSGWNP 258
Qy 241 OPSCEEMTCLTPYPNGIYTPHRIKRIIDDEIRYCKNGFYPATRSPVSKCTITGWPAP 300
Db 259 OPSCEEMTCLTPYPNGIYTPHRIKRIIDDEIRYCKNGFYPATRSPVSKCTITGWPAP 318
Qy 301 RCLSKPCDFPOFKHGRLYYESSRRPYFPVPIGKEYSYCYDNGFTTPSOYWDYLRCTVNG 360
Db 319 RCLSKPCDFPOFKHGRLYYESSRRPYFPVPIGKEYSYCYDNGFTTPSOYWDYLRCTVNG 378
Qy 361 WEPEVPCLRQCFIHYVEYGESSYWRORYIEGOSAKVCHSGYSLPNGODTYCTENGWSP 420
Db 379 WEPEVPCLRQCFIHYVEYGESSYWRORYIEGOSAKVCHSGYSLPNGODTYCTENGWSP 438
Qy 421 PPKCVRIK 428
Db 439 PPKCVRIK 446

RESULT 2
ID Q9NU87 PRELIMINARY; PRT; 1172 AA.
AC Q9NU87;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DJ177p10.1.1 (H FACTOR 1 (COMPLEMENT) ISOFORM 1).
GN HFL
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bird C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL049744; CAB70597.1; -.
DR HSSP: P08603; IHFH.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 19.
DR SMART: SM00032; CCP; 19.
SQ SEQUENCE 1172 AA; 132087 MW; 8F5B954C4B4FA454 CRC64;

Query Match 69.8%; Score 1714; DB 4; Length 1172;
Best Local Similarity 67.5%; pred. No. 2e-149;
Matches 289; Conservative 47; Mismatches 92; Indels 0; Gaps 0;

Qy 1 EDCKGPPPRENSEILSGSWSEQLYSEGTOATYKCRPGYRTLGTTVYCKNGEWYPSNPSR 60
Db 19 EDCNELPPRANTEILTSWSWSDQTYPEGTQATYKCRPGYRSIGNVIMVCRKGEWALNPLR 78
Qy 61 ICRKRPCGHPGDTFPGFRLAVGSEFEFGAKVYVTCDEGYQLLGEIDYRECDADGWTNDI 120
Db 79 KCQRRPCGHPGDTFPGFTTGGNVFYGKAVYTCDEGYQLLGEINRYRECDTGWTNDI 138
Qy 121 PICEVVKCLPVTLENGRIVSGAAEPDQYFGVGVFECNSGFKIEGQKEMHCSENGLW 180
Db 139 PICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVFVCSNGYKIEGDEMHCSDDGFW 198
Qy 181 SNEKPCQVEISCLPPRVENGDIYLLKPVYKENERFOYKCKOGFYKRGDAVCTGSGWNP 240
Db 199 SKEPKVCIESCKSPDVINGSPISQKIYKENERFOYKCNMGVEYSERGDAVCTESGWRP 258
Qy 241 OPSCEEMTCLTPYPNGIYTPHRIKRIIDDEIRYCKNGFYPATRSPVSKCTITGWPAP 300
Db 259 LPCEEKSNDNYPNGDYSPLRKIKHTGDEITVOCRNNGFYPATRGNAKTSTGTWPAP 318
Qy 301 RCLSKPCDFPOFKHGRLYYESSRRPYFPVPIGKEYSYCYDNGFTTPSOYWDYLRCTVNG 360
Db 319 RCLSKPCDFPOFKHGRLYYESSRRPYFPVPIGKEYSYCYDNGFTTPSOYWDYLRCTVNG 378
```

```
Db 319 RCLSKPCDYPDIKHGLYHENMRRPYFPVAVGKYYSYCYDEHFETPSGSYWDHIHCTQDG 378
Qy 361 WEPEVPCLRQCFIHYVEYGESSYWRORYIEGOSAKVCHSGYSLPNGODTYCTENGWSP 420
Db 379 WSPAVPCLRKCYFPYLENGYNONHGRKRFVQGSIDVACHPGYALPQAOTTVTCMENGWSP 438
Qy 421 PPKCVRIK 428
Db 439 TPCIRVK 446

RESULT 3
ID Q28085 PRELIMINARY; PRT; 669 AA.
AC Q28085;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CCP MODULES 3-12. WITH PARTS OF CCP 2 AND 13 (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=36202005; PubMed=8615824;
RA Soames C.J., Day A.J., Sim R.B.;
RT "Prediction from sequence comparisons of residues of factor H involved
in the interaction with complement component C3b.";
RL Blochem. J. 315:523-531(1996).
DR EMBL: X98697; CAA67257.1; -.
DR HSSP: P10998; 1VVD.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 11.
DR SMART: SM00032; CCP; 11.
FT NON_TER 1
FT NON_TER 669
SQ SEQUENCE 669 AA; 75683 MW; D0D9DB30EE747AC2 CRC64;

Query Match 50.6%; Score 1242; DB 6; Length 669;
Best Local Similarity 58.4%; pred. No. 3.7e-106;
Matches 206; Conservative 57; Mismatches 88; Indels 2; Gaps 1;

Qy 76 GSFRLAVGSEFEFGAKVYVTCDEGYQLLGEIDYRECDADGWTNDIPICEVVKCLPVTLE 135
Db 3 GSPHLAEGNQFEYGAHVYTCDEGYQVGMENFRCDTGWTNDIPICEVVKCLPVTLE 62
Qy 136 NGRIVSGAAEPDQYFGVGVFECNSGFKIEGQKEMHCSENGLWSNEKPCQVEISCLPP 195
Db 63 NGKIFSDALEPDQYFGVGVFECNSGYMLDGPQKIHCSAGGVSAETPKCVEIFCKPP 122
Qy 196 RVENGDDGIYLLKPVYKENERFOYKCKOGFYKRGDAVCTGSGWNPQSCHEMTCLTPYP 255
Db 123 VILNGQAVLPRATYKQNERVOYRCAAGFEYQGRDGTVCYKSGWTPAPTICETICPPRIP 182
Qy 256 NGIYTPHRIKRIIDDEIRYCKNGFYPATRSPVSKCTITGWPAPRCSLKPCDFPOFKHG 315
Db 183 NGVYRPELSKRGQDKTYIECKKGFFPIRGTDTCTRDGWPVPRCAWKPCSYPIVKHG 242
Qy 316 RLYYEESRRPYFPVPIGKEYSYCYDNGFTTPSOYWDYLRCTVNGWEPEVPCLRQCFIHY 375
Db 243 RLYY--SYRGYFPAVNVQFVYSCDHHFVPPSQRSWDHLACTAGWSPEEPCLRQCFIHY 300
Qy 376 VEYGESSYWRORYIEGOSAKVCHSGYSLPNGODTYCTENGWSPPPKCVRIK 428
Db 301 LENGHNRREKYLQGETVRVHCYEGYSLQNDQNTMTCTESGWSPPPCIRVK 353

RESULT 4
Q61407
```

```
ID Q61407 PRELIMINARY; PRT; 452 AA.
AC Q61407;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE COMPLEMENT FACTOR H-RELATED PROTEIN (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA Chaplin D.D.;
RL Submitted (APR-1990) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90153969; PubMed=1689298;
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA Chaplin D.D.;
RT "Identification and sequence analysis of 4 complement factor H-related
RT transcripts in mouse liver.";
RL J. Biol. Chem. 265:3193-3201(1990).
DR EMBL; M29010; AAA37415.1; -
DR HSSP; P08603; 1HFI.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 7.
DR SMART; SM00032; CCP; 7.
DR NON_TER 452 452
SQ SEQUENCE 452 AA; 51602 MW; 2B697A4FFC6E13CA CRC64;

Query Match 33.1%; Score 813; DB 11; Length 452;
Best Local Similarity 71.0%; Pred. No. 9.2e-67;
Matches 137; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

QY 236 SGWNPQPSCEEMTCLTPYIPNGIYTPHRIKHRIIDDEIRYCKNGFYPATRSPVSKCTITG 295
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15 TAWLSTAKEEKTCSPPYILNGIYTPHRIIHRSDDEIRYECNYGFPYVTGSTVSKCTPTG 74

QY 296 WIPAPRCLKPCDFPOFKHGRLYEESRRPYPVPPIGKEYSYCDNGFTTPSQSYWDYLR 355
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 75 WIPVPRCTLKPCFEPFKYGRLYEESLRPNFVSGNKNYSYKCDNGFSPPSGYSWDYLR 134

QY 356 CTVNGWEPVPCLRQCFHYHYVEGESSYWRRIYEGOSAKVOCHSGYSLPNGQDTYYCTE 415
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 135 CTAQGWEPVPCVRKCVFHYVENGDSYWEKIYVQOSLKVOCYNGYSLQNGQDTMTCTE 194

QY 416 NGWSPPPKCVRIK 428
| | | | | | | | | |
Db 195 NGWSPPPKCIK 207

RESULT 5
Q61405 PRELIMINARY; PRT; 303 AA.
AC Q61405;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE COMPLEMENT FACTOR H-RELATED PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90153969; PubMed=1689298;
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA Chaplin D.D.;
RT "Identification and sequence analysis of 4 complement factor H-related
RT transcripts in mouse liver.";
RL J. Biol. Chem. 265:3193-3201(1990).
```

```
DR EMBL; M29007; AAA37413.1; -
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
SQ SEQUENCE 303 AA; 34498 MW; 51C66E50906F4C24 CRC64;

Query Match 33.0%; Score 811; DB 11; Length 303;
Best Local Similarity 70.8%; Pred. No. 8.5e-67;
Matches 136; Conservative 19; Mismatches 37; Indels 0; Gaps 0;

QY 236 SGWNPQPSCEEMTCLTPYIPNGIYTPHRIKHRIIDDEIRYCKNGFYPATRSPVSKCTITG 295
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15 TAWLSTAKEEKTCSPPYILNGIYTPHRIIHRSDDEIRYECNYGFPYVTGSTVSKCTPTG 74

QY 296 WIPAPRCLKPCDFPOFKHGRLYEESRRPYPVPPIGKEYSYCDNGFTTPSQSYWDYLR 355
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 75 WIPVPRCTLKPCFEPFKYGRLYEESLRPNFVSGNKNYSYKCDNGFSPPSGYSWDYLR 134

QY 356 CTVNGWEPVPCLRQCFHYHYVEGESSYWRRIYEGOSAKVOCHSGYSLPNGQDTYYCTE 415
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 135 CTAQGWEPVPCVRKCVFHYVENGDSYWEKIYVQOSLKVOCYNGYSLQNGQDTMTCTE 194

QY 416 NGWSPPPKCVRI 427
| | | | | | | | | |
Db 195 NGWSPPPKCI 206

RESULT 6
Q91275 PRELIMINARY; PRT; 1053 AA.
AC Q91275;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE COMPLEMENT REGULATORY PLASMA PROTEIN.
OS Parabrax nebulifer (barred sand bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Seranilidae; Paralabrax.
OX NCBI_TaxID=30873;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=LIVER;
RC MEDLINE=94318039; PubMed=8042982;
RA Dajmen A., Kaidoh T., Zipfel P.F., Gigli I.;
RT "Cloning and characterization of a cDNA representing a putative
RT complement-regulatory plasma protein from barred sand bass (Parabrax
RT nebulifer).";
RL Biochem. J. 301:391-397(1994).
DR EMBL; L21703; AAA92556.1; -
DR HSSP; P08603; 1HFI.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 16.
DR SMART; SM00032; CCP; 16.
SQ SEQUENCE 1053 AA; 117597 MW; F27E32C3AD76D5D3 CRC64;

Query Match 23.2%; Score 569.5; DB 13; Length 1053;
Best Local Similarity 26.7%; Pred. No. 8.2e-44;
Matches 148; Conservative 60; Mismatches 169; Indels 177; Gaps 19;

QY 21 EQLYSEGTOATYKCRPGVRLTGLTIYKVCCKNGEWPNFSRICKRPPCHGPGDTPGSRFL 80
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 47 EASTPGGRQVRVGCNVGYS--GFFKLVCVEGKWETRGAK--CQPRSCGHPGDAQFADPHL 102

QY 81 AVGEFEFGAKVYVTCDEGYQLLGEIDYRECDAGTWNIDPICEVVKCLPYTELENGRIV 140
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 103 AEGNDFVFGSKVYVTCQGYQMSRINRVRCVAGWDGVVPCVCSQQC-PLIHVDNNVQV 161

QY 141 SGAAEPDQEEYFGOVVRFECNSGPKI-EGQKEMHCSENGLSNEKPKQVCISCLPPRVEN 199
```

```
Db 162 IGPGE---EATGNYVRECKRSEILDSGSELYCDERGDWSPVKKKAITCAIPPIEN 218
QY 200 GGIYLYKPVYKNERFOYKCKGFFYKRGDAVCTGSG-----WNPOPSCEEMTCL----- 250
Db 219 GNVPGAIREYKENDVLYHCEADRAFKHIDR-PTCIKQGIKAESWPTPLCESIKRCLTIMD 277
QY 251 -TPYIP-----NGIYT----- 260
Db 278 GTRYEPAYRNLFSPGETLKVICARTSWISTPOETSVVTTQDNGEWSIRPTCOEVRCSNR 337
QY 261 -PHRI-----KHRIIDELRYECKNGFYPATRSPVSKCTITGWIPAP----- 300
Db 338 RPEHVDSDVRSWERYTLDNTRYKCKRGYKRTGGVWTATCGRNGMMPNPLCEVCKVTCCKE 397
QY 301 -----RCSLKPCCDFPQFKH 314
Db 398 NIODAVIVCTDKQYLNKAIYACGEGNRRITLTCGNGWSGDRKCTVKPCPLPKDP 457
QY 315 GRLYEESRRPFPVPIGKEYSYCDNGFTTTPSQSYWDYLRCTVNGWEPEVPCLOCI-- 372
Db 458 NGFF-----RGPY-----TGRVLYYTKDGYKLFTEGHWAEAKVDGVW-----PELTTCISN 505
QY 373 -----PHYVEYGESSYQWRRYIEGSAKVQCHSGYSLPNSQDTYYCTENGW----- 418
Db 506 TTCGKFPPEIPNAEVI---RYPEVQTVQVICNQGS--TQANSFSCENGWLLYGLSPDQ 560
QY 419 -----SPPPK 423
Db 561 ICTLRADVCGPPPE 574

RESULT 7
Q61408 ID Q61408 PRELIMINARY; PRT; 808 AA.
AC Q61408:
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE COMPLEMENT FACTOR H-RELATED PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90153969; PubMed=1689298;
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA Chaplin D.D.;
RT "Identification and sequence analysis of 4 complement factor H-related
transcripts in mouse liver."
RL J. Biol. Chem. 265:3193-3201(1990).
DR EMBL; M29009; AAA37416.1; -.
DR HSP; P08603; IHCC.
DR InterPro; IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00084; sush1; 13.
DR SMART; SM00032; CCP; 13.
SQ SEQUENCE 808 AA; 91654 MW; 6FD97D53CE74DF6D CRC64;
```

```
Query Match 23.1%; Score 566; DB 11; Length 808;
Best Local Similarity 64.5%; Pred. No. 1.2e-43;
Matches 91; Conservative 23; Mismatches 27; Indels 0; Gaps 0;
```

```
QY 288 VSKCTITGWIPAPRSLKPCDFPQFKHRLYYEESRRPFPVPIGKEYSYCDNGFTTTPS 347
Db 9 LSNILLTAWLSTAKEVSKCEFPQFKYGRLYFEELRPNFPVIGNKYSYKCDNGFSPPS 68
QY 348 OSYWDYLRCTVNGWEPEVPCLOCIIFHYVEYGESSYQWRRYIEGSAKVQCHSGYSLPNG 407
Db 69 GLFWDYLRCTVQGMKPEVPCVRKCVFHYVYENGFEAYWEKIIYVQGSGLKVCQYNGYSLQNG 128
QY 408 QTYTYCTENGWSPPPKCVRIK 428
```

```
Db 129 QDTMTCTENGWSPPPKCIRIK 149

RESULT 8
Q99JAI ID Q99JAI PRELIMINARY; PRT; 555 AA.
AC Q99JAI:
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE C4BP ALPHA-CHAIN PRECURSOR.
GN C4BPA.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=EPIDIDYMS, AND LIVER;
RX MEDLINE=21154058; PubMed=11254714;
RA Nonaka M.I., Wang G., Mori T., Okada H., Nonaka M.;
RT "Novel androgen-dependent promoters direct expression of the C4b-
binding protein alpha-chain gene in epididymis."
RL J. Immunol. 166:4570-4577(2001).
DR EMBL; AB049465; BAB39737.1; -.
DR EMBL; AB049466; BAB39738.1; -.
DR EMBL; AB049467; BAB39739.1; -.
DR HSP; P10998; LVVD.
DR InterPro; IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00084; sush1; 8.
DR SMART; SM00032; CCP; 8.
KW Signal.
FT SIGNAL 1 13 POTENTIAL.
FT CHAIN 14 555 POTENTIAL.
SQ SEQUENCE 555 AA; 61628 MW; 7AFA5462AFC7B6AC CRC64;
```

```
Query Match 19.5%; Score 478; DB 11; Length 555;
Best Local Similarity 27.2%; Pred. No. 1e-35;
Matches 140; Conservative 73; Mismatches 178; Indels 124; Gaps 27;

QY 5 GPPPRENSILSGWSEQLYSEGTQATYKCRPGY-RTLGTIVKVK-NGEWPVSNFSRIC 62
Db 16 GPPVLFATPAKELNQTETTTNLKTYCRPGYIRHSDQTLTCKVNGW---RYEVFC 72
QY 63 RRPCHPGDTPGSPRLAVSGSEFEFGAKVYTCDSGYQLLG-----EIDYRECDADGW 116
Db 73 SKKQCRNPGDLPHGT--IEVKTDLFLGSKIEFSGFNLVGPPTTSYCEIHDKGVD--W 127
QY 117 TNDIPICEVVKCLPVTLENGRIVSGAEPDQBYFYGVVVFECNSGFKIEGKEMHCS- 175
Db 128 SYVFPICEIILKCRSPPDISNGK-HSGAVE--DLTYTGSSVYTYSCDPSYLLGNPSISCTV 184
QY 176 ---ENGLWNEKPOQVEISCLPRVE-----NGDGIYLYKPVYKNERFOYKCKGFFYKE 227
Db 185 VAKTVGWSPPVCKKVICRQPIVOYANLISGFG---PIYTKDTIMFSCOKGFVLKG 240
QY 228 RGDVAC-TGSGWNPQS-CEEMTCL-TPYIPNG-----IYTPHRIKHRIDDEI 272
Db 241 SSLIRCEADNNWNPNPFPVCEPNCSVDIPDIPAYWDSYRRPRKGLYSPGTV-----F 293
QY 273 RYCKNGFYPATRSPVSKCTIT-----GWIPAPRSLKPCDFPQFKHGR----- 316
Db 294 KTSCHSGYVPAH---ESTVTVCQDFKSPFRGCKKVCPCAPEVNGNSMRADSYSTD 350
QY 317 -----LYEESRRPYF-----VPYIG-----KE----- 334
Db 351 CPFSYNNIFQYRCDSRQYITSTTCQADGTWKPRVFCGQACHVPEPEIAHGRYKEGYLSA 410
QY 335 ---YSYCDNGFTTQSGSYWDYLRCTVNGWEPEVP-CLROCIFHYVEYGESSYQWRRYIEG 391
Db 411 LSYVVECDGDTLVGQ---NTITCKNSLSSEAPQCAKQCLPKIENGKLSVDKPOYIEP 467
```

```
Qy 392 QSAKVQCHSGYSLPNGQDTYYCTENG-WSP-PPKC 424
   : : | | : : | | : : | | : : | | : : | |
Db 468 ENVTIHCDSGFKL-EGSPSITCSEKGTWHGVPKC 501

RESULT 9
ID Q63135 PRELIMINARY: PRT: 559 AA.
AC Q63135;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE COMPLEMENT REGULATORY PROTEIN.
GN CRRY.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=96006570; PubMed=7590969;
RA Ouligg R.J., Lo C.F., Alexander J.J., Sneed A.E., Moxley G. III;
RT "Molecular characterization of rat Crry: widespread distribution of
RT two alternative forms of Crry mRNA.";
RL Immunogenetics 42:362-367(1995).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Dohi N., Sakurada C., Nonaka M., Okada N., Okada H.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; L36532; AAA91821.1; -.
DR EMBL; BA42115; BAA22548.1; -.
DR HSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 7.
DR SMART; SM00032; CCP; 7.
DR SEQUENCE 559 AA; 61680 MW; 29E10F6A21DB9B6E CRC64;

Query Match 19.0%; Score 466; DB 11; Length 559;
Best Local Similarity 26.6%; Pred. No. 1.3e-34;
Matches 122; Conservative 91; Mismatches 190; Indels 56; Gaps 26;

Qy 3 CKGPPPRENSILSGSWSEQLYSGTQATYKCRGY--RTLGTIVKVKNGEWPSPNSR 60
   | | : : : : | | : : | | : : | | : : | |
Db 38 CPAPPLFPYAKPIPT-DESTFPVGTSLKYECRPGYIKRQFSITCEV--NSVW--TSPOD 92

Qy 61 ICRKRCGHPGDTFPGSPRLAAGSEFERGAKVVTCDGSGYLLGEIDYRECDAD---GWT 117
   | : : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
Db 93 VCIRKQETPLDPQNGI--VHVNTDIRGSSITTCNBYRLIGSSAMCIISQSVAMD 150

Qy 118 NDIPICEVVKCLPYTELENGRIVSGAEPDQE-YFVGQVRFECNSG-----FKIEGQK 170
   | | | | : : | | : : | | : : | | : : | | : : | | : : | |
Db 151 AEAPICESIPCEIIPSPINGDFES----PNREDFHGMVVTYQCNTDARGKKLENLVCEP 206

Qy 171 EMHCS-----ENGLSNKEPKQCVELS-CLPPRVNODGIYL-KPVYKENERQYCKQGFV 224
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 207 SIHCTSIDGQGVWGGPPPPQCIELNKTTPPHVENAVIVSKNLSFLSLDMVFEFCQDGF 266

Qy 225 YKRGDAVCTG-SQWNPO-PSCEMTCLTPY---IPNG-IYTPHRIKHRIIDEIRYECKN 278
   | | | | : : | | | | : : | | : : | | : : | | : : | | : :
Db 267 MKGSSVYCRSLNWEQLPSCFVKVSGAGELGPNHGVFPQNL--OLGAKVTFVCNT 324

Qy 279 GFYPATSPVSKNTITG----W-IPAPRCSLKPCDFQFKHGRLYYESRRPYFPVPIGK 333
   | | | | : : | | : : | | : : | | : : | | : : | | : : | |
Db 325 G-YLKGKNSSHCVLDGVESINWSSVPCEQVICKLPQDMSFGFKGLQMKDDY---YGD 380

Qy 334 EYSYCDNGFTTSPQSWDYLRCTVN-GWEPEVP-CLRQCIFHYVE----YGESSYQWR 387
   | : : : | : : | : : | : : | : : | : : | : : | : : | : :
Db 381 NVALECEDGTYTLEGG---OSQCQSDASWDPLPKCVSQVICKLPQDMSGFQKGLQMKD 437
```

```
Qy 388 YIEGSAKVQCHSGYSLPNGQDTYYCTENGWSP-PPKV 425
   | | : : | | : : | | : : | | : : | | : : | |
Db 438 YYYGDNVALECEDGYTLEGGSSQSQSQSDASMDPLPKCV 476

RESULT 10
ID Q99254 PRELIMINARY: PRT: 679 AA.
AC Q99254;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE COMPLEMENT RECEPTOR TYPE 2 (CR2) (FRAGMENT).
GN CR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=90229754; PubMed=2139460;
RA Kurtz C.B., O'Toole E., Christensen S.M., Weis J.H.;
RT "The murine complement receptor gene family. IV. Alternative splicing
RT of Cr2 gene transcripts predicts two distinct gene products that share
RT homologous domains with both human CR2 and CR1.";
RL J. Immunol. 144:3581-3591(1990).
[2]
RN [2]
RP SEQUENCE OF 21-367 FROM N.A.
RX MEDLINE=95105691; PubMed=7528766;
RA Kim Y.O., Kinoshita T., Molina H., Hourcade D., Seya T., Wagner L.M.,
RA Holers V.M.;
RT "Mouse complement regulatory protein Crry/p65 uses the specific
RT mechanisms of both human decay-accelerating factor and membrane
RT cofactor protein.";
RL J. Exp. Med. 181:151-159(1995).
DR EMBL; U17128; AAA78271.1; -.
DR EMBL; U17123; AAA78271.1; JOINED.
DR EMBL; U17124; AAA78271.1; JOINED.
DR EMBL; U17125; AAA78271.1; JOINED.
DR EMBL; U17126; AAA78271.1; JOINED.
DR EMBL; U17127; AAA78271.1; JOINED.
DR EMBL; M36470; AAA37449.1; -.
DR HSP; P10998; 1VVD.
DR MGD; MGI:98489; Cr2.
DR InterPro; IPR002396; Selectin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 10.
DR PRINTS; PR00343; SELECTIN.
DR SMART; SM00032; CCP; 10.
DR NON_TER 679 679
SQ SEQUENCE 679 AA; 74916 MW; 52FC00FDCED20CDC CRC64;

Query Match 18.5%; Score 453; DB 11; Length 679;
Best Local Similarity 28.6%; Pred. No. 2.7e-33;
Matches 126; Conservative 73; Mismatches 178; Indels 64; Gaps 27;

Qy 24 YSEGTQATYKCRPGVYRTLTGTVKVKNGEWPSPNSRICKRKPCHGPGDTPFGSFLAVG 83
   : : | | | | | | | | : : | : : | : : | : : | : : | : :
Db 42 FAIGTWYKCRPGVFRKSFITCLETSKW--SDAQQCFKRPCKMNPQEPHSGSVHINTG 99

Qy 84 SEFEFGAKVVTCDGSGYLLGEIDYRECDAGQ---WTNDIPICEVVKCLPYTELENGRIV 140
   | | | | : : | | : : | | : : | | : : | | : : | | : : | |
Db 100 --IEFGSTITYSCNGYRLIGDSSATCIVSDNTVMWMDMPLCESIPCSPPAISNGDFY 157

Qy 141 SGAAEPDDEYFVGQVRFECNSG-----FKIEGQKEMHCSEN----GLWSNEKQCV-E 189
   | : : : | : : | : : | : : | : : | : : | : : | : : | : :
Db 158 SSSRD---SFFYGMVVTYCHTKNRKRLFDLVGKSIYCTSKDNQGVGWNPPQCIPR 214

Qy 190 ISCLPPRVENG---DGIYKPKVYKENERQYCKQGFYKRGDAVC-TGSGWN-Q-PSC 244
   | : : | | | | | | : : | : : | : : | : : | : : | : :
Db 215 VKCPMEIENGLVESG---FKHSFFLNDVTIFRKSGFTMKGSRIAMQPNKWSPLPTC 272
```


DR	InterPro; IPR001424; SOD_CU_ZN.
DR	InterPro; IPR000436; Sushi_SCR_CCP.
DR	InterPro; IPR000834; Zn_carbopept.
DR	Pfam; PF00084; sushi; 30.
DR	SMART; SM00032; CCP; 30.
DR	PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR	PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW	Receptor.
SQ	SEQUENCE 2039 AA; 223603 MW; B82FCB1LC6B16635 CRC64;

Query Match	17.8%; Score 437; DB 4; Length 2039;
Best Local Similarity	27.7%; Pred. No. 3.2e-31;
Matches 132; Conservative	63; Mismatches 149; Indels 132; Gaps 31;

Qy	27	GTOATYKCRPGYRGLGIIVKVCKNGEWPSPNSRICRKRCPCGHGDDTDFGFRILAVGEF	86
Dd	66	GTYLNECRPGYGRPFESIICLKSNVMTGAKDR--CRKSKSRNPDPVNGVMHVTK--I	121
Qy	87	EFGAKVVVTCDGGYQLLGEIDYRECDADG----WTNDIPICEVKC-LPVTELENGRTVS	141
Dd	122	QFGSQIKYSCTKGRLIGS--SSATCIISGDVIWNDETPICDRIPCGLPPT-ITNGDFTS	179
Qy	142	GAAPEDQEYYVGQVVRFECSNG-----FKIEGOKEMHCSEN-----GLWSNEKPQC-V	190
Dd	180	TNRE---NFHYGVSVTVTRCNPGSGRKVFELVGPESIVCTSDNQDVQVLSGPAQCII	236
Qy	191	SCLPFRVNG---DGIIYLYKYVENERFYKKQGFFYKERGDAVCTG-SGNWPK-PSC	244
Dd	237	KCTPNVENGILVSDN---RSLFLSNVEVERFCQGFVMMKPPRRVKCOALNKWEPLPSC	293
Qy	245	EEMTCLPYIPNGIYTPHRIKHRIID----DIRYECKNGFYPATRSPVSKCTITG-WIP-	298
Dd	294	SRV-CQPP--PDVLHAERTORDKNFSPGQEVFYSCPG-YDLRGAAAMRCTPOGDHSPA	349
Qy	299	APRCSLPACD--FPQFKHGRIYYEESRRYPVP--PTGKEYSYVDNGFTT--PSQSYMD	352
Dd	350	APTCEVSCDDFMGOLLNGRV-----LPPNLQLGAKYDFVCDEGFLKLGSASV--	399
Qy	353	YLRTCVAAG---WBEPVPCLRQCIFHYVEGYESSYWQRRYEGQSAAVQCHSGYSLPN--	406
Dd	400	--CVLAGMESLNSSVPVCEQ-----;-----IFCPSPVIPNGR	431
Qy	407	-----GDTTYCTE-----NG-W-SPPPKC	424
Dd	432	HTGKPLEVFPFGKAVNYTCDPHPDRGTSTFDLIGESTIRCTSDPCNGCWSSPAPRC	487

RESULT	15
ID	Q16744
AC	Q16744 PRELIMINARY; PRT; 2489 AA.
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE	COMPLEMENT RECEPTOR 1.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=94065175; PubMed=8245463;
RA	Vik D.P., Wong W.W.;
RT	"Structure of the gene for the F allele of complement receptor type 1
RT	and sequence of the coding region unique to the S allele.";
RL	J. Immunol. 151:6214-6224(1993).
RL	[2]
RP	SEQUENCE FROM N.A.
RA	Vik D.P., Wong W.W.;
RL	Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
DR	EMBL; L17418; AAB60695.1; -
DR	EMBL; L17390; AAB60695.1; JOINED.

DR	EMBL; L17399; AAB60695.1; JOINED.	
DR	EMBL; L17409; AAB60695.1; JOINED.	
DR	EMBL; L17419; AAB60695.1; JOINED.	
DR	EMBL; L17420; AAB60695.1; JOINED.	
DR	EMBL; L17421; AAB60695.1; JOINED.	
DR	EMBL; L17422; AAB60695.1; JOINED.	
DR	EMBL; L17423; AAB60695.1; JOINED.	
DR	EMBL; L17424; AAB60695.1; JOINED.	
DR	EMBL; L17425; AAB60695.1; JOINED.	
DR	EMBL; L17426; AAB60695.1; JOINED.	
DR	EMBL; L17427; AAB60695.1; JOINED.	
DR	EMBL; L17428; AAB60695.1; JOINED.	
DR	EMBL; L17429; AAB60695.1; JOINED.	
DR	EMBL; L17430; AAB60695.1; JOINED.	
DR	EMBL; L17391; AAB60695.1; JOINED.	
DR	EMBL; L17392; AAB60695.1; JOINED.	
DR	EMBL; L17393; AAB60695.1; JOINED.	
DR	EMBL; L17394; AAB60695.1; JOINED.	
DR	EMBL; L17395; AAB60695.1; JOINED.	
DR	EMBL; L17396; AAB60695.1; JOINED.	
DR	EMBL; L17397; AAB60695.1; JOINED.	
DR	EMBL; L17398; AAB60695.1; JOINED.	
DR	EMBL; L17400; AAB60695.1; JOINED.	
DR	EMBL; L17401; AAB60695.1; JOINED.	
DR	EMBL; L17402; AAB60695.1; JOINED.	
DR	EMBL; L17403; AAB60695.1; JOINED.	
DR	EMBL; L17404; AAB60695.1; JOINED.	
DR	EMBL; L17405; AAB60695.1; JOINED.	
DR	EMBL; L17406; AAB60695.1; JOINED.	
DR	EMBL; L17407; AAB60695.1; JOINED.	
DR	EMBL; L17408; AAB60695.1; JOINED.	
DR	EMBL; L17410; AAB60695.1; JOINED.	
DR	EMBL; L17411; AAB60695.1; JOINED.	
DR	EMBL; L17412; AAB60695.1; JOINED.	
DR	EMBL; L17413; AAB60695.1; JOINED.	
DR	EMBL; L17414; AAB60695.1; JOINED.	
DR	EMBL; L17415; AAB60695.1; JOINED.	
DR	EMBL; L17416; AAB60695.1; JOINED.	
DR	EMBL; L17417; AAB60695.1; JOINED.	
DR	HSP; P08603; I HFI.	
DR	InterPro; IPR001424; SOD_CU_ZN.	
DR	InterPro; IPR000436; Sushi_SCR_CCP.	
DR	InterPro; IPR000834; Zn_carbopept.	
DR	Pfam; PF00084; sushi; 37.	
DR	SMART; SM00032; CCP; 37.	
DR	PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_3.	
DR	PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.	
KW	Receptor.	
QY	SEQUENCE. 2489 AA; 272846 MW; CEE11B53F2B4FAF6 CRC64;	
Query Match 17.8%; Score 437; DB 4; Length 2489;		
Best Local Similarity 27.7%; Pred. No. 4.2e-11;		
Matches 132; Conservative 63; Mismatches 149; Indels 132; Gaps 31;		
QY	27 GTQATYKRCRGYRTLGIIVKVCNGEWPNSPRICRKRPGCHGDPFGSFRLAGVSEF 86	
DB	66 GTYLNVECRPGYSGRPSIICLNKSNVMTGAKDR--CRRKSCRNPPDPVNGMVHVIKG--I 121	
QY	87 EFGAKVVTYCDGQYQLGEIDYRECDADG----WTNDIPICEVVKC-LPTVTELENGRIVS 141	
DB	122 QFGSIKYSCTKGYRLIGS--SSATCIISGDTVIMDNETPIDCRIPCLGLPT-ITNGDFIS 179	
QY	142 GAAPEDQYVFGVVRPECNSSG-----FKIEGQKEMHCSEN-----GLWSEKPOC-VEI 190	
DB	180 TNRE---NFHYGSVVYTRCNPNGSGGRKRVFLVGEPSITCNSDDGVINGSPAPCCLIPN 236	
QY	191 SCLPPRVENG-----DGIYLAPYKENERFYKQKQGFYKRGKGVKRGDAVCTG-SGNPQ-PS 244	
DB	237 KCTPPNVENGILYSDN--RSLFSLNVEVFEFCQPFVYKMGPRRVKQCALNKWEPELPSC 293	
QY	245 EEWTLCTPYTPNGIYTPHRIKTRID----DEIYEKNGGYPATRSPVSKCTITG-WIP- 298	

Query Match	17.8%	Score 437;	DB 4;	Length 2489;
Best Local Similarity	27.7%;	Pred. No. 4.2e-31;		
Matches 132;	Conservative 63;	Mismatches 149;	Indels 132;	Gaps 31;
QY	27	GTQATYKRCGCGYRTLGTIVKCKNGEWPVNSPRICBPKCGHGGDPFGSFRFLAVCSEF	86	
DB	66	GTYNVCEGCGYSGRPSIICLNKSVMTGADR--CRKSKRNPPDPVNGVMVWIKG--I	121	
QY	87	EFGAKVYVTCDEGYQLLGEIDYRECDADG----WTNDIPICEVVKC-LPYTELENGRIVS	141	
DB	122	QFGSQIKYSCTKGRLIGS--SATCIISGDTVIWNETPIDCRIPGLPPT-ITNGDFIS	179	
QY	142	GAAPEDQYFFGQVVRPECHSG-----FKLEGCKEHCSEN-----GLWSNEHPQC-VEI	190	
DB	180	TNRE---NFHYGSVTVYRCNPGSGGRKVFELVGPESITCTNSDDGVINGSPACIIPN	236	
QY	191	SCLPVRPVENG-----DGIYLKPVYKENERFOYKCKQGQVYKERGDAVCTG--SGNPO--PSC	244	
DB	237	KCTPPNVENGILVSDN---RSLSELNVEVFECQGFVYMGKPRRVKQALNKKWEPELPC	293	
QY	245	EWMTCLTPYIPNGIYTPHRIKTRID----DEIRYCKNGFYFATRSPVSKCTITG-WIP--	298	

Db 294 SRV-COPP--PDVLHAERTQDKDNFSPGOEVFYSCBPG-YDLRGAASMRCTPOGDWSPA 349
Qy 299 APRCSLKPCD--FPQFXHGRLYYEESSRRPYFPV--PICKEYSYCDNGFTT--PSQSYWD 352
Db 350 APTCEVAKSCDDFMGQLLNGRV-----LFPVNLQLCAKVDVFCDEGQFLKGSSASY-- 399
Qy 353 YLRCTVNG---WEPEVPCLRQCIFHYVEYGESSYWQPRYIEGOSAKVQCHSGYSLPN-- 406
Db 400 ---CVLAGMESLWNSVPVCEQ-----IFCPSPVPVTPNGR 431
Qy 407 -----GQDTYYCTE-----NG-W-SPPPKC 424
Db 432 HTGKPLEVFPFGKAVNYTCDPHDPDRGTSFDLIGESTIRCTSDPOGNGVWSSPAPRC 487

Search completed: August 29, 2002, 15:11:44
Job time: 605 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:03:40 ; Search time 75.55 Seconds
(without alignments)
629.247 Million cell updates/sec

Title: US-09-316-163-14
Perfect score: 2454
Sequence: 1 EDCKGPPRENSILSGWS.....DTYYCTENGWSPPKVCVRK 428

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	990	40.3	240 18 AAW39154	Human partial Comp
2	918	37.4	216 18 AAW39155	Clone PRRB9FH410 C
3	544.5	22.2	578 20 AAY09065	Human complement f
4	481.5	19.6	581 12 AAR13490	Human C4 binding p
5	438	17.8	543 13 AAR28547	CRI-4 (52S, 53S, 5
6	437	17.8	453 20 AAY55757	Human CRI protein
7	437	17.8	778 19 AAW73147	Amino acid sequenc
8	437	17.8	1930 19 AAW45899	Human complement r
9	437	17.8	2039 20 AAY55751	Human C3b/C4b rece
10	437	17.8	2039 22 ABG00287	Novel human diagno
11	437	17.8	2044 22 ABB11782	Human CRI protein

12	437	17.8	2044	22	AAW39224	Human polypeptide
13	437	17.8	2044	22	AAW41010	Human polypeptide
14	437	17.8	2317	10	AAP92219	CRI protein. Homo
15	434	17.7	496	20	AAV55752	Human CRI protein
16	434	17.7	543	13	AAR28557	CRI-4 (99H, 103E)
17	434	17.7	1537	12	AAR11982	Partial human comp
18	434	17.7	1929	22	ABG00103	Novel human diagno
19	434	17.7	2039	14	AAR36743	CRI-4 (266-274 KLK
20	432	17.6	543	13	AAR28570	Human complement t
21	432	17.6	2039	12	AAR11810	Human CRI protein
22	431	17.6	450	20	AAV55754	Human CRI protein
23	430	17.5	543	13	AAR28560	CRI-4 (114S) analo
24	429	17.5	543	13	AAR28548	CRI-4 (57V, 59K) a
25	428	17.4	543	13	AAR28550	CRI-4 (64K) analog
26	428	17.4	543	13	AAR28553	CRI-4 (85R, 87N) a
27	428	17.4	543	13	AAR28558	CRI-4 (109N, 110A,
28	428	17.4	543	13	AAR28565	CRI-4 (121Q) analo
29	428	17.4	543	13	AAR28571	CRI-4 (364-367 NAA
30	427	17.4	543	13	AAR28545	CRI-4 (37Y) analog
31	427	17.4	543	13	AAR28552	CRI-4 (78T, 79D) a
32	426	17.4	543	13	AAR28555	CRI-4 (92T) analog
33	426	17.4	543	13	AAR28563	CRI-4 (117P) analo
34	426	17.4	543	13	AAR28567	CRI-4 (318-321 RNP
35	425	17.3	543	13	AAR28549	CRI-4 (64K, 65T) a
36	425	17.3	543	13	AAR28551	CRI-4 (65T) analog
37	425	17.3	543	13	AAR28562	CRI-4 (116K) analo
38	425	17.3	543	13	AAR28568	CRI-4 (347T, 349Y)
39	424	17.3	543	13	AAR28569	CRI-4 (369-376 STK
40	424	17.3	645	21	AAW53125	Maceca mulatta rha
41	423	17.2	543	13	AAR28556	CRI-4 (94H) analog
42	423	17.2	543	13	AAR28564	CRI-4 (116K, 117P)
43	422	17.2	543	13	AAR28561	CRI-4 (115T) analo
44	421	17.2	543	13	AAR28554	CRI-4 (92T, 94H) a
45	420	17.1	543	13	AAR28544	CRI-4 (35E) analog

ALIGNMENTS

RESULT 1
AAW39154
ID AAW39154 standard; Protein; 240 AA.
XX
AC AAW39154;
XX
DT 27-APR-1998 (first entry)
XX
DE Human partial Complement factor H protein fragment 1.
XX
DE Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medicament; modulator.
XX
OS Homo sapiens.
XX
PN WO9738136-A1.
XX
PD 16-OCT-1997.
XX
PF 09-APR-1997; 97WO-US05710.
XX
PR 06-MAR-1997; 97US-0812481.
PR 09-APR-1996; 96US-0015083.
PR 09-APR-1996; 96US-0630048.
PR 06-MAR-1997; 97US-0038614.
XX
PA (BARD-) BARD DIAGNOSTIC SCI INC.
XX
XX Enfield DL, Hass GM, Kinders RJ;
XX
XX WPI: 1997-512742/47.
DR N-PSDB; AAV02790.
XX
PT Treating or screening for cancer, e.g. renal or urogenital cancer -

PT by modulating or detecting tumour associated human complement Factor
 PT H related antigen, or nucleic acid encoding it
 XX
 PS Example 6B; Fig 6B; 104pp; English.
 XX

CC This partial protein sequence represents a region of the human
 CC tumour-associated complement factor H (CFH). This sequence is used
 CC in the identification of complement factor H related proteins and
 CC antigens isolated from clone PRBB9FH410 (see AAW39155). The detection of
 CC such proteins and a CFH antigens can be used in screening or for the
 CC treatment of renal or urogenital cancer, e.g. bladder, cervical or
 CC prostate cancer. Agents that may modulate this antigen could be used in
 CC the manufacture of a medicament for the treatment of a tumour cell.
 XX
 SQ Sequence 240 AA;

Query Match 40.3%; Score 990; DB 18; Length 240;
 Best Local Similarity 71.7%; Pred. No. 1e-67;
 Matches 172; Conservative 21; Mismatches 47; Indels 0; Gaps 0;

QY 78 FLAVGSEPEFGAKVYTCDEGYQLGEIDYRECDADGWTNDIPICEVVKCLPVTLENG 137
 DB 1 flltgnvfeygkavtycneqyllgeinrecdtgdwtdipicevkvclpvtapeng 60
 QY 138 RIVSGAAEPDQYYFGVVFECNSGFKIEGQKEMHCSENGLSNEKPCQVEISCLPPRV 197
 DB 61 kiwssamepdreyhfgavrfvcnsgyklegdeemhcsddgfwskpkcveiscspdv 120
 QY 198 ENGDGYLKPVKYKRGDAVCTGSGWNPQPSCEMTCLTPYPNG 257
 DB 121 ingdspisqliykenerfykcmgyeysergdavctesgwrplpsceekscdnpypng 180
 QY 258 IYTPHRIKHIDDEIRVECKNGFYATRSVPVSKCTITGWIPAPRCSLKPCDFQFKHGL 317
 DB 181 dysplrikhrtgdeitygcrgnfyatrgntaktctgtgwipaprcctikpcdydkhgg1 240

RESULT 2
 AAW39155
 ID AAW39155 standard; Protein; 216 AA.
 XX
 AC AAW39155;
 XX

DT 27-APR-1998 (first entry)
 DE
 DE Clone PRBB9FH410 CFH related protein fragment.
 XX
 KW Complement factor H; tumour associated antigen; renal cancer;
 KW urogenital cancer; medicament; modulator.
 XX
 OS Synthetic.
 XX
 PN WO9738136-A1.
 XX
 PD 16-OCT-1997.
 XX

PF 09-APR-1997; 97WO-US05710.
 XX
 PR 06-MAR-1997; 97US-0812481.
 PR 09-APR-1996; 96US-0015083.
 PR 09-APR-1996; 96US-0630048.
 PR 06-MAR-1997; 97US-0038614.
 XX
 PA (BARD-) BARD DIAGNOSTIC SCI INC.
 XX
 PI Enfield DL, Hass GM, Kinders RJ;
 XX
 DR WPI; 1997-512742/47.
 DR N-PSDB; AAW02791.
 XX

PT Treating or screening for cancer, e.g. renal or urogenital cancer -
 PT by modulating or detecting tumour associated human complement Factor

PT H related antigen, or nucleic acid encoding it
 XX
 PS Example 6B; Fig 6B; 104pp; English.
 XX

CC This partial protein is found in clone PRBB9FH410 and represents a
 CC complement factor H related protein with homology to a region of the
 CC human tumour-associated complement factor H (CFH). The detection of this
 CC protein and a CFH antigen can be used in screening or for the treatment
 CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.
 CC Agents that may modulate this antigen could be used in the manufacture of
 CC a medicament for the treatment of a tumour cell.
 XX
 SQ Sequence 216 AA;

Query Match 37.4%; Score 918; DB 18; Length 216;
 Best Local Similarity 73.1%; Pred. No. 2.8e-62;
 Matches 158; Conservative 19; Mismatches 39; Indels 0; Gaps 0;

QY 95 TCDEGYQLGEIDYRECDADGWTNDIPICEVVKCLPVTLENGRIVSGAAEPDQYYFGQ 154
 DB 1 tcnegvqllgeinrecdtdgwtndipicevkvclpvtapengkiwssamepdreyhfgq 60
 QY 155 VVFECSNGFKIEGQKEMHCSENGLSNEKPCQVEISCLPPRVENGDIYLPVKYKENER 214
 DB 61 avrfvcnsgyklegdeemhcsddgfwgkpkcveiscspdvingspiskliykener 120
 QY 215 FQYKCKQGFYKRGDAVCTGSGWNPQPSCEMTCLTPYPNGIYTPHRIKHIDDEIRY 274
 DB 121 fgykcmgyeysergdavctesgwrplpsceekscdnpypngdysplrikhrtgdeity 180
 QY 275 ECKNGFYATRSVPVSKCTITGWIPAPRCSLKPCDFP 310
 DB 181 qcrngfyatrgntaktctgtgwipaprcctikpcdydp 216

RESULT 3
 AAY09065
 ID AAY09065 standard; Protein; 578 AA.
 XX
 AC AAY09065;
 XX

DT 06-JUL-1999 (first entry)
 DE
 DE Human complement factor H homolog protein.
 XX
 KW Human complement factor H; immunological mechanism; complement reaction;
 KW gene therapy; immune stimulation; haematopoiesis regulation; chemotactic;
 KW tissue growth activity; anti-inflammatory; tumour inhibition;
 KW secretory signal.
 XX
 OS Homo sapiens.
 XX
 PN WO9918200-A1.
 XX
 PD 15-APR-1999.
 XX

PF 02-OCT-1998; 98WO-JP04448.
 XX
 PR 06-OCT-1997; 97JP-0272837.
 XX
 PA (PROT-) PROTEGENE INC.
 PA (SAGA) SAGAMI CHEM RES CENT.
 XX
 PI Kato S, Sekine S;
 XX
 DR WPI; 1999-264019/22.
 DR N-PSDB; AAX34737.
 XX

PT Human proteins with secretory signal sequences and nucleotide
 PT sequences, useful in control of proliferation and differentiation of
 PT cells
 XX

Claims 1; Page 55-58; 71pp; English.

This DNA encodes a protein having homology to human complement factor H, which plays a role in the immunological mechanism involving the complement reaction. The protein can also be used as an antigen for preparing antibodies against the protein. The cDNA can be used as a probe for gene diagnosis and the gene for gene therapy, as well as for large-scale expression of the protein. The protein may also have immune stimulating or suppressing activity, haematopoiesis regulating activity, tissue growth activity, activin/inhibin activity, anti-inflammatory activity, tumour inhibition activity, chemotactic/chemokinetic activity, receptor/ligand activity, etc. The protein is identified by the presence of a hydrophobic N-terminal secretory signal region, knowledge of the protein function is not required, as in e.g. methods of expression cloning.

Sequence 578 AA:

Query Match 22.2%; Score 544.5; DB 20; Length 578;

Best Local Similarity 32.3%; Pred No. 2.9e-33;

Matches 136; Conservative 55; Mismatches 147; Indels 83; Gaps 24;

QY 48 CKNGEWPSPSRICKRRCGHP-----GDPFGSFR-----LAVGSEFFGAKVVYTCDE 98

Db 15 cangqev-----kpcdfpeiqhggllyksrlrlyfpaaagsys-----yycdq 58

QY 99 GYQL-LGEI-DYRECDADGWINDIPCEVVKLPV-----TELENGRIVSGAAEPDOEYY 151

Db 59 nvtptsgsywdyihctqdgwstvp-----clrtcsksadvelengfi-----sesssiyi 108

QY 152 FGQVRFECNSGFKI---EGQEHMCSENGLSNEKPOCWEISCLPPRVEN---GDGIY 204

Db 109 lneetqncpkyatadgnssgsitclqng-wstq-picikf-cdmpvfensraksngmw 165

QY 205 LKPVYKENERFOYKQKGF--VYKERGDA-VCTGSWNPPQSC---EEMTCLTPYIPNG- 257

Db 166 ----fklhdtldecydygessyngntdsivcgedgwhlptcynssescgppppisngd 221

QY 258 -----IYTPRIKHRIIDDETRYECKNGFYATPSVSKCITGWIPAPRC-SLUKPCDF 309

Db 222 ttfspqkvyilpw-----srveygcqs-yyelqgskyyvtcsngdwsepprcismkpcsf 273

QY 310 POFKIGRLYYEESRRYPVPPIGKEYSYCDNGFTTPSOSYWDYLRCTVNGWEPEVPCLR 369

Db 274 peiqhghlyventrpfvatgqsyysydcndqnfvtpsgyswdyihctqdgwlpvtvpcir 333

QY 370 QCIFH--VVEYGESSYWMORRYIEGOSAKVQCHSGYSLPNQDQ--TYICTENGWSPPPKCV 425

Db 334 tcsksadieiengfissesssiylinkeiykckpgyatadgnssgsitclqngwsaqpicl 393

QY 426 R 426

Db 394 k 394

RESULT 4

ID AAR13490 standard; Protein; 581 AA.

XX AAR13490;

AC AAR13490;

XX 30-OCT-1991 (first entry)

XX Human C4 binding protein.

XX C4bp; monomer; complement protein; pJOD.C4bp.3; SCR;

KW short consensus repeat.

XX Homo sapiens.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..32

FT Protein /label= signal_peptide

FT 33..581

FT /label= C4bp

FT 33..93

FT /label= SCR8

FT 94..155

FT /label= SCR7

FT 156..219

FT /label= SCR6

FT 220..279

FT /label= SCR5

FT 280..345

FT /label= SCR4

FT 346..406

FT /label= SCR3

FT 407..464

FT /label= SCR2

FT 465..523

FT /label= SCR1

FT 524..581

FT /label= C4bp_core

FT /note= "responsible for multimer assembly"

FT Disulfide-bond 34..80

FT /note= "intradomain"

FT Disulfide-bond 65..92

FT /note= "intradomain"

XX WO9111461-A.

PN 08-AUG-1991.

XX 28-JAN-1991; 91WO-US00567.

XX 26-JAN-1990; 90US-0470888.

XX (BIOG-) BIOGEN INC.

XX Pasek MP, Winkler G, Liu TR;

XX WPI; 1991-252613/34.

XX N-PSDB; AAK13242.

XX New C4 binding protein fusion proteins and DNA encoding them -

XX comprise assemblies of C4bp monomers linked to functional moiety,

XX e.g. AZT, useful as delivery vehicles in diagnosis and therapy

XX Example 1; Fig 1; 105pp; English.

XX This sequence was deduced from human hepatocyte (Hep G2) cDNA

XX obtained following PCR amplification. The protein is a monomer

XX containing 8 SCRs. Each SCR forms a looped domain due to the

XX presence of two intradomain disulphide bonds (only the disulphide

XX bonds of SCR8 are labelled in the Features Table). Within each SCR,

XX the first cysteine residue bonds with the third and the second

XX cysteine residue bonds with the fourth. This secondary structure is

XX responsible for the conformational flexibility of the C4bp monomer.

XX The invention covers fusion proteins in which the monomer sequence,

XX or a specified subfragment of it, e.g. having 5, 4, 3 or 1 SCR(s)

XX is fused to the C-terminal of a protein such as a viral receptor,

XX cell ligand, a bacterial, viral or parasitic immunogen, enzyme,

XX cytokine, toxin, etc. See also AAK13243-51.

XX Sequence 581 AA;

XX Query Match 19.6%; Score 481.5; DB 12; Length 581;

XX Best Local Similarity 26.5%; Pred. No. 1.8e-28;

XX Matches 134; Conservative 78; Mismatches 195; Indels 99; Gaps 25;

QY 2 DCKGPPPRENSEILSGWSEQLYSEGTOATYKRCPGY-RTLGTIVKVC-KNGEWPSPNS 59

Db 33 ncgppptlsfaapmditltetrfktgttkytcclpgyvrshstqtlctnsdgywnt--- 90

QY 60 RICRRPCGHPGDTFGSFRSLAVGSEFEFGAKVVTCTDEGYQLLGEIDYRECDAD---G 115
 Db 91 -fcykrchrpdelng--qveikdlfsqlefcscsegffligttsr-cevqdrvg 146
 QY 116 WTNDIPICEVVKCLPVTLENGRIVSGAAEPDQYFFGQVVRFECSGFKIEGQKEMHCS 175
 Db 147 wshpqpceivkckppdrrng-----hsgeenfyaygsvtyscdprfslighasict 202
 QY 176 -EN--GLWSNEKPCQVEISCLPPRVENGDIY-LKPVYKENERFOYKCKQGFVYKERGD 230
 Db 203 venetigwrrpsptcekitcrkpdshgemvsgfpiynykdtivfkckqgfvf--rqs 260
 QY 231 AV--C-TSGWNPO-PSCEMTCLT-PYIPNGIYT--PHRIK---HRIDDEIRYECKNGF 280
 Db 261 svihcdadskwnpspacepncinldiphaswetyprtkedvvyvgtviryrcpgy 320
 QY 281 YPATRSPVSKCITG--WIPAPRCSLKPCDFPQFKHGLRYEESRRP----- 325
 Db 321 kptdeptvicqnlrwtptyggcealccepeklmgteitqrkrpranhvfyfygdel 380
 QY 326 -----YPPVPIG-----KEYSYVYCDN 341
 Db 381 fschetsrfsaicqgdgtwsptrpcgdicnfpkiahgkyqssysffkeeliyecdk 440
 QY 342 GFTTFSQSYWDYLRCTVNGWEPEVP-CLRQCIFHYVEGESSYWRRYIEGOSAKVQCHS 400
 Db 441 gyilvgqa---kiscsyshwsapqckalcrkpelvngrlsvdkdyvpepntvtdqds 497
 QY 401 GYSLPNGDYYCTEN-GWSP-PRC 424
 Db 498 gygvv-gpqsitcsgrtwtvpevpkc 522

RESULT 5

AAR28547

ID AAR28547 standard; peptide; 543 AA.

XX AC AAR28547;

XX DT 19-MAR-1993 (first entry)

XX DE CR1-4 (52S, 53S, 54P) analogue.

XX KW short consensus repeat; regulator of complement activation;

XX OS Homo sapiens.

XX FT Key

XX FT Region

XX FT Location/Qualifiers

XX FT 1..60

XX FT /label= SCR-1

XX FT 61..122

XX FT /label= SCR-2

XX FT 451..510

XX FT /label= SCR-8

XX FT 511..543

XX FT /label= SCR-9

XX FT /note= "TRUNCATED"

XX FT Misc-difference 52

XX FT /note= "Thr substituted by Ser (SCR-8)"

XX FT Misc-difference 53

XX FT /note= "Gly substituted by Ser (SCR-8)"

XX FT Misc-difference 54

XX FT /note= "Ala substituted by Pro (SCR-8)"

XX PN EP512733-A.

XX PD 11-NOV-1992.

XX PF 28-APR-1992; 92EP-0303826.

XX PR 03-MAY-1991; 91US-0695514.

XX XX

PA (UNIW) UNIV WASHINGTON.

XX Atkinson JP, Hourcade D, Krych M;

XX WPI; 1992-375009/46.

XX Complement activity regulator protein analogues - useful for

XX treating auto-immune diseases, to suppress transplant rejection,

XX for diagnosis etc.

XX Claim 11; Fig 2 and R11810; 23pp; English.

XX The cDNA clone designated CR1-4 was described in J.Exp.Med.(1988)

XX 168:1255-1270. It encodes the first 8 and a half amino terminal

XX SCRs of CR1. The invention concerns analogues of "regulator of

XX complement activation" proteins or truncated, hybrid or recombinant

XX forms of them. CR1-4 is a preferred truncated form and a number of

XX specified substitution variants of it are claimed in which certain

XX positions in SCR-1 which have been identified as important for the

XX degree of C3b- and C4b-binding are substituted by amino acids for

XX the corresponding positions in SCR-8. The specification does not

XX contain the CR1-4 sequence; the sequence given here was constructed

XX from the full-length CR1 amino acid sequence having GENESEQO

XX accession number AAR11810 and descriptions in the disclosure.

XX Sequence 543 AA;

SQ

Query Match 17.8%; Score 438; DB 13; Length 543;

Best Local Similarity 27.9%; Pred. No. 3.5e-25;

Matches 133; Conservative 63; Mismatches 148; Indels 132; Gaps 31;

QY 27 GTOATYKCRPGYRTLTGTVKCKNGEWPSPNPSRICRKRKCGHPGDTFGSFRSLAVGSEF 86

Db 25 gtylnyecrgpygrpfslieclknsvw--spkdrckrkrscnppdvngmvhvikg--l 80

QY 87 EFGAKVYVTCDEGYQLLGEIDYRECDAG-----WTNDIPICEVVKC-LPVTLENGRIVS 141

Db 81 qfgsqikysctkgyrligs--ssatclisgdtvldnetpicdrpcglpplp-itngdfis 138

QY 142 GAAEPDQYFFGQVVRFECSG-----PKIEGQKEMHCSEN-----GLWSNEKPCQ-VEI 190

Db 139 ttre---nfhygsvvtyrcnpsggrkvfelvgepslyctsnddqvgiwsppqclpn 195

QY 191 SCLPPRVENG-----DGIYLYKPVYKENERFOYKCKQGFVYKERGDVCTG-SGNWPO-PS 244

Db 196 kctppnvengllvsdn---rsifsnvevfrcqpfvmkgprvrkckqalnkwepelpsc 252

QY 245 EMTCLTPYIPNGIYTPHRIKHRID----DEIRYECKNGFYPATRSPVSKCTITG-WIP- 298

Db 253 srv-cqpp--pdvlhaertqrkdnfspqgevfyscepq-ydlrgaasmrctpqgdwspa 308

QY 299 APRCSLKPCD--FPQFKHGLRYEESRRPYFPV--PIGKEYSYVCDNGFTT--PSQSYWD 352

Db 309 aptcevkscddfmqgllngrv-----lfpvniqlgakvdfvdegfqlkgssasy-- 358

QY 353 YLRCTVNG----WEPEVPCLRQCIFHYVEGESSYWRRYIEGQSAKVCHSGYSUPN-- 406

Db 359 ---cvlagmeslwnssvpvceq-----GQDTYYCTE-----NG-W-SPPPKC 424

QY 407 -----GQDTYYCTE-----NG-W-SPPPKC 424

Db 391 htgkplevfpgkavnytcddpdpdrgtsfdlligestirctsdpgqngvswspaprc 446

RESULT 6

AAY55757

ID AAY55757 standard; Protein; 453 AA.

XX AC AAY55757;

XX DT 22-FEB-2000 (first entry)

XX XX

[illegible]

```

QY 142 GAAPDQEYVFGVVFECNSG-----FKLEGKEMHCSEN---GLWSNEKPOC-VEI 190
Db 139 tnre---nflygsvvtvrcnpgsggrkvfvlvgepslyctsnddqvglwsgpapqclipn 195
QY 191 SCLPPRVENG-----DGIYLPVYKENERFOYKCKQGfVYKERGDVACTG-SGNWPO-PSG 244
Db 196 kctppnvengllvsdn--rslfslnevvefricqpgfvmkgprvrkcaqlnkwepepisc 252
QY 245 EEMTCLTPYIPNGIYTPHRIKRID----DEIRYCKNGFYYPATRSPVSKCTITG-WIP- 298
Db 253 srv-cqpp--pdvlihaertqrdkntfspgqevfyscepg-ydlrgaasmrctpggdwsa 308
QY 299 APRCSLAKPCD--PQPKHGRLYEESSRRPFPV--PIGKEYSYICNGFTT--PSOSYWD 352
Db 309 aptcevkscdfmgqlingrv-----lfvnlqlgakvdcvdegfqlkgssasy-- 358
QY 353 YLRCTVNG----WEPEVPCLQRQICFIHYVEYGESSYQRRYIEQSAKVOCHSGYSLPN-- 406
Db 359 --cvlagmeslnssvpvceq-----ifcspvpvlpngr 390
QY 407 -----GQDTYYCTE-----NG-W-SPPPKC 424
Db 391 htgkplevfpgkavnytdcpdhrgtsfdaligestirdspqgngvwssaprc 446

RESULT 9
AAAY55751
ID AAAY55751 standard; Protein; 2039 AA.
XX
XX AAAY55751;
XX
XX
XX 22-FEB-2000 (first entry)
XX DT
XX Human C3b/C4b receptor (CR1) protein.
XX

```


PN WO200153312-A1.
XX 26-JUL-2001.
PD 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR N-PSDB; AAI60166.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX Example 2; SEQ ID NO 5941; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAI38642-AAI42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX Sequence 2044 AA;

Query Match 17.8%; Score 437; DB 22; Length 2044;
Best Local Similarity 27.7%; Pred. No. 2.2e-24;
Matches 132; Conservative 63; Mismatches 149; Indels 132; Gaps 31;
QY 27 GTOATYKRCPCGYRTLTGTTIVKCKNGEWPVNSPTRICRKRPGCHPDTPFGSFRVAVGSEF 86
DB 71 gtylnyrcpysgrpsfiiclknsvtgakdr--crkscrnppdpvngmvhvk-g-i 126
QY 87 EFGAKVYVTCDEGYQLLGEIDYRECDADG----WTNDIPICEVVKC-LPVTLENGRIVS 141
DB 127 qfsgqikysctkgryllgs--ssatciilsgdvtviwdnetpicdrpcglpbt-ingdfis 184
QY 142 GAAPDEQYFYGQVVRRECNSG-----FKIEGQKEMHCSEN----GLWSNEKPOC-VEI 190
DB 185 tnre---nfhygsvvtvrcnpgsggrkvfelvgepslyctsnddqvgiwsqpapqciup 241
QY 191 SCLPVRVENG----DGIYLPKPYKNERFQYKCKGQFVYKERGDAVCTG-SGWNQ-PSC 244
DB 242 kctppnvengilvsdn---rsfslnevfeircpgfvmkgprvrkcaqlnkwepeipsc 298
QY 245 EEMTCLTPYIPNGIYTPHRIKHRID----DEIRYECKNGFYPAATRSVPVKCTITG-WIP- 298
DB 299 srv-cqpo--pdvllaertqrdkdnfsgpgqevfyscepg-ydlrgaasmrctpgdgwsa 354

QY 299 APRCSLKPCD--FPQFKHGRLLYYESRRPYFPV--PIGKEYSYCDNGFTT--PSOSYWD 352
DB 355 aptcevkscddfmqqlngrv-----lfpvniqlgakvdfvcdgflkgssasy-- 404
QY 353 YLRCTVNG----WEPEVCLROCIEFHYVEYGESSYWMORRYIEGQSAKVOCHSGYSLPN-- 406
DB 405 ---cvlagmeslwnssvpvceq-----ifcspppvlpnqr 436
QY 407 -----GQDTYYCTE-----NG-W-SPPKRC 424
DB 437 htgkplevfpgkavnyctdphdrgtsfdligestirctsdpgngvswsspaprc 492
RESULT 14
AAP92219
ID AAP92219 standard; protein; 2317 AA.
XX AC AAP92219;
XX DT 22-FEB-1990 (first entry)
XX DE CRI protein.
XX KW Complement; cofactor.
XX OS Homo sapiens (human).
XX FH Key Location/Qualifiers
FT Peptide 10..50
FT /label= signal_peptide
PN WO8909220-A.
PD 05-OCT-1989.
XX 31-MAR-1989; 89WO-US01358.
XX 01-APR-1988; 88US-0176532.
XX (TCEL) T CELL SCIENCES INC.
XX (UJJO) THE JOHNS HOPKINS UNIVERSITY.
XX (BRIG) THE BRIGHAM AND WOMEN'S HOSPITAL.
XX Fearon DT, Klickstein LB, Wong W, Carson G, Concino MF, Makrides SC;
XX WPI: 1989-309498/42.
XX N-PSDB: AAN91477.
XX New nucleic acid sequences encoding new CRI protein - and its fragment,
PT for diagnosis and control of complement-related immune defects,
PT inflammation, myocardial infarct, etc
XX Claim 1; Fig 1; 191pp; English.
XX This is full-length CRI protein, and shortened forms are new, lacking
CC the transmembrane region. The proteins and fragments bind C3b and/or
CC C4b, have cofactor activity and inhibit C3 and C5 convertase activity.
CC In the sequence, x-untranslated region. This has 7 short consensus
CC repeats (CR) of 4 long direct homologous repeats (LHR) (see fig. 10 in
CC patent). Deletion mutants, eg LHR-A for binding of C4b, and LHR-C
CC for C3b binding, be made, and LHR-B and -C can detect CR2 sequences.
CC They are useful in diagnosing and treating immune disorders, and prevent
CC perfusion injury.
XX Sequence 2317 AA;
SQ
Query Match 17.8%; Score 437; DB 10; Length 2317;
Best Local Similarity 27.7%; Pred. No. 2.6e-24;
Matches 132; Conservative 63; Mismatches 149; Indels 132; Gaps 31;
QY 27 GTOATYKRCPCGYRTLTGTTIVKCKNGEWPVNSPTRICRKRPGCHPDTPFGSFRVAVGSEF 86
DB 71 gtylnyrcpysgrpsfiiclknsvtgakdr--crkscrnppdpvngmvhvk-g-i 126
QY 87 EFGAKVYVTCDEGYQLLGEIDYRECDADG----WTNDIPICEVVKC-LPVTLENGRIVS 141
DB 127 qfsgqikysctkgryllgs--ssatciilsgdvtviwdnetpicdrpcglpbt-ingdfis 184
QY 142 GAAPDEQYFYGQVVRRECNSG-----FKIEGQKEMHCSEN----GLWSNEKPOC-VEI 190
DB 185 tnre---nfhygsvvtvrcnpgsggrkvfelvgepslyctsnddqvgiwsqpapqciup 241
QY 191 SCLPVRVENG----DGIYLPKPYKNERFQYKCKGQFVYKERGDAVCTG-SGWNQ-PSC 244
DB 242 kctppnvengilvsdn---rsfslnevfeircpgfvmkgprvrkcaqlnkwepeipsc 298
QY 245 EEMTCLTPYIPNGIYTPHRIKHRID----DEIRYECKNGFYPAATRSVPVKCTITG-WIP- 298
DB 299 srv-cqpo--pdvllaertqrdkdnfsgpgqevfyscepg-ydlrgaasmrctpgdgwsa 354

